# REPORT RESUMES

ED 003 552 56

AN ANALYSIS OF THE BEHAVIORAL PROCESSES INVOLVED IN SELF-INSTRUCTION WITH TEACHING MACHINES.

BY- SKINNER, B.F. HOLLAND, JAMES G.

HARVARD UNIV., CAMBRIDGE, MASS.

REPORT NUMBER NDEA-VIIA-191 PUB DATE 64
GRANT OEG-71-31-0370-051-3
EDRS PRICE MF-$0.36 HC-$8.16 204P.

DESCRIPTORS- *TEACHING MACHINES, *BEHAVIOR PATTERNS, *VERBAL LEARNING, *AUTOINSTRUCTIONAL AIDS, LINEAR PROGRAMING, PROGRAMED INSTRUCTION, CAMBRIDGE, MASSACHUSETTS

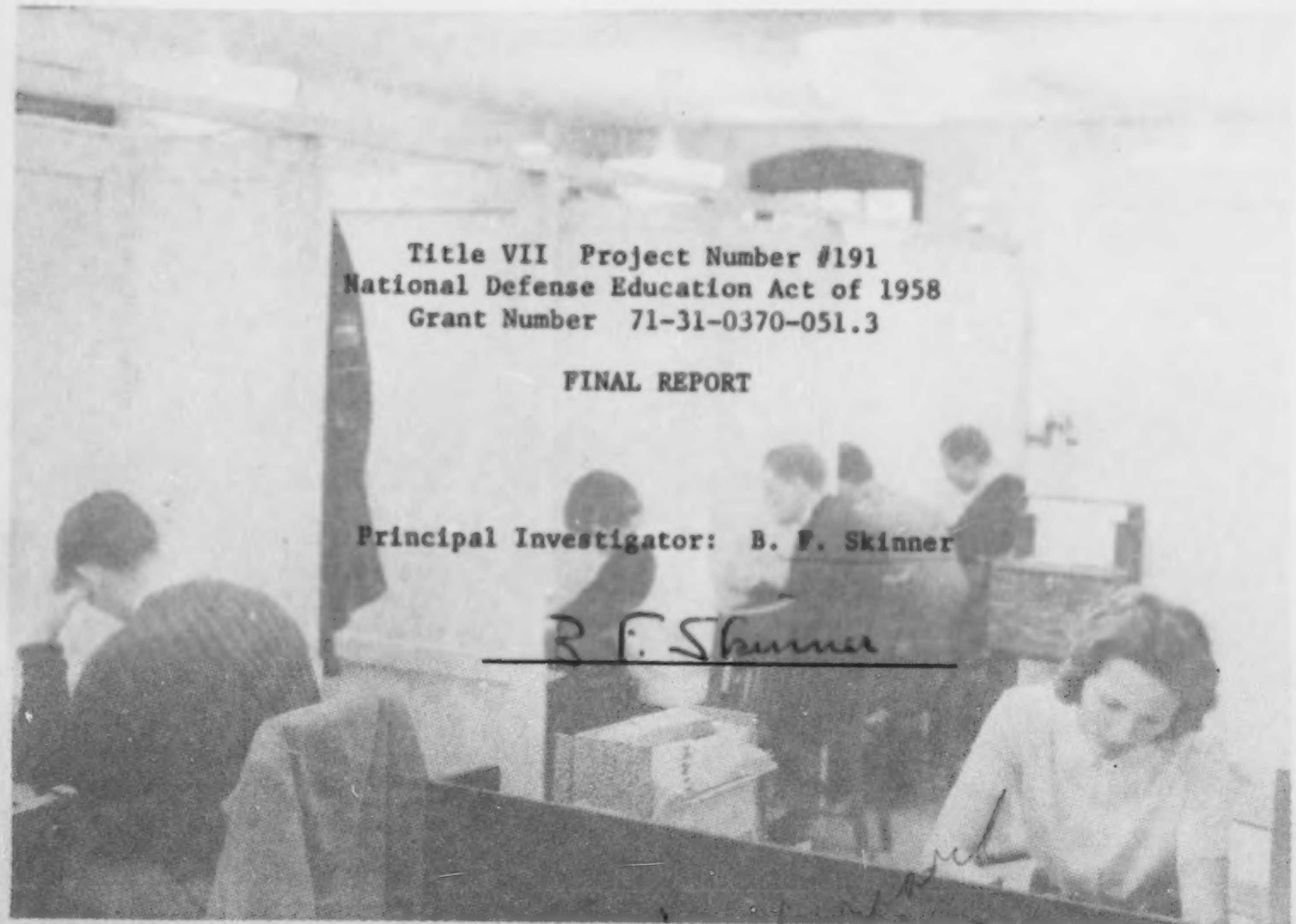
THIS COLLECTION OF PAPERS CONSTITUTES THE FINAL REPORT OF A PROJECT DEVOTED TO AN ANALYSIS OF THE BEHAVIORAL PROCESSES UNDERLYING PROGRAMED INSTRUCTION. THE PAPERS ARE GROUPED UNDER THREE HEADINGS--(1) "PROGRAMING RESEARCH," (2) "BASIC SKILLS--RATIONALE AND PROCEDURE," AND (3) "BASIC SKILLS--SPECIFIC SKILLS." THE SUMMARY WRITTEN BY THE PRINCIPAL INVESTIGATOR IS GIVEN IN A PAPER TITLED "REFLECTIONS ON A DECADE OF TEACHING MACHINES." MANY OF THE PAPERS HAVE BEEN PUBLISHED IN PROFESSIONAL JOURNALS. (AL)

ED003552

191

# An Analysis of the Behavioral Processes Involved in Self-Instruction with Teaching Machines

Editors: James G. Holland
B. F. Skinner



Title VII Project Number #191
National Defense Education Act of 1958
Grant Number 71-31-0370-051.3

## FINAL REPORT

Principal Investigator: B. F. Skinner

B. F. Skinner

Harvard Univ

The Research Reported Herein Was
Supported by a Grant from the
U.S. Department of Health, Education, and Welfare
Office of Education

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

# An Analysis of the Behavioral Processes Involved in Self-Instruction with Teaching Machines

Editors: James G. Holland
B. F. Skinner

Title VII Project Number #191
National Defense Education Act of 1958
Grant Number 71-31-0370-051.3

FINAL REPORT

Principal Investigator: B. F. Skinner

B. F. Skinner

The Research Reported Herein Was
Supported by a Grant from the
U.S. Department of Health, Education, and Welfare
Office of Education

# TABLE OF CONTENTS

i Preface

Verbal Programming Research

1 Response Contingencies in Teaching-Machine Programs
James G. Holland

9 Evaluating Teaching Machines and Programs
James G. Holland

19 The Influence of Repetition of Incorrectly-Answered Items in a Teaching-Machine Program
James G. Holland and Douglas Porter

23 A Note on Machines in a Technology of Education
William C. Holz and John Robinson

25 Variably Blurred Prompting in the Analysis of Paired-Associate Learning
Matthew Israel

31 The Incorporation of a PSSC Film into a Program
Philip G. Schrag and James G. Holland

Basic Skills: Rationale and Procedure

35 Why We Need Teaching Machines
B. F. Skinner

57 New Directions in Teaching-Machine Research
James G. Holland

67 A Multiple-Choice Visual Discrimination Apparatus
Wells Hively

71 Programming Stimuli in Matching to Sample
Wells Hively

91 Parametric Experiments with a Matching to Sample Program
Wells Hively

Basic Skills: Specific Skills

105 Systematic Instruction in the Attainment of Right-Left Form Concepts in Young and Retarded Children
Sidney Bijou

137 An Investigation of Key-Tone Matching with Children and Adults
Edward Maltzman

161 Teaching a Skill of Data Interpretation by Programmed Instruction
Vernon Stone

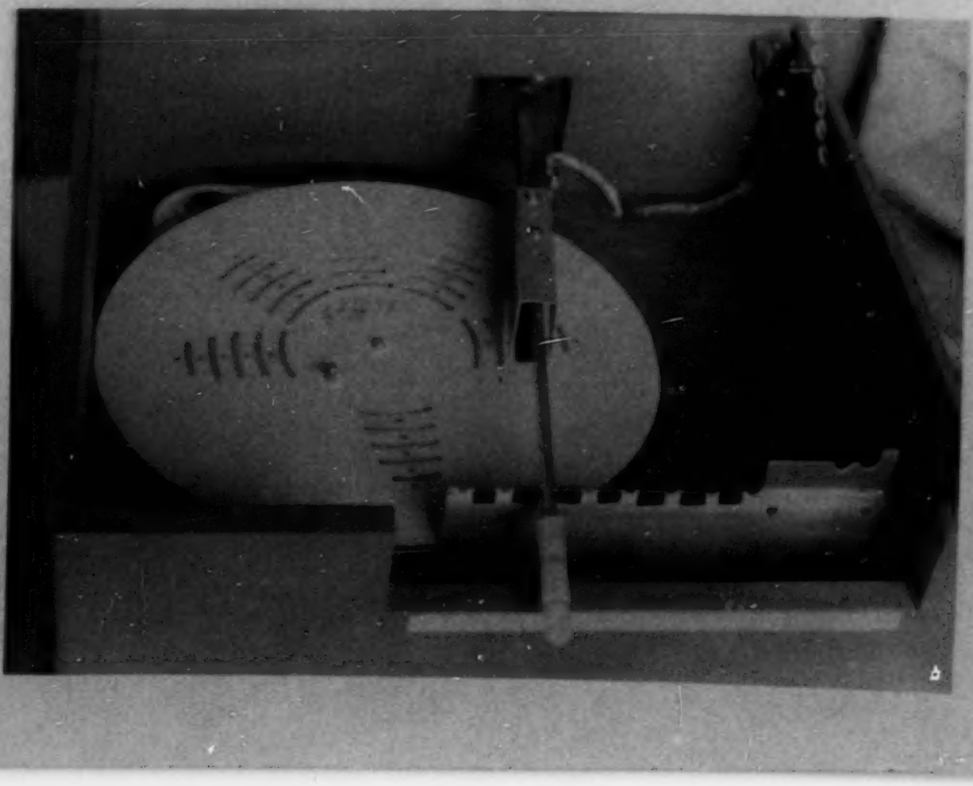
165 Attempts to Program Materials to Teach Certain Elements of a Phonic Repertoire
Wells Hively and Helen M. Popp

177 Visual Discrimination of Alphabet Letters
Helen M. Popp

SUMMARY

183 Reflections on a Decade of Teaching Machines
B. F. Skinner

# VERBAL PROGRAMMING RESEARCH



# PREFACE

This collection of papers constitutes the final report of research supported by Title VII of P.L. 191, National Defense Education Act of 1958. The collection consists of the tangible products of the project and contains only the research reports, judged by the authors, to be appropriate in form and quality for publication in professional journals. Some have been published and are reproduced here as they appeared in their original journal publication. Most of the remaining are either in press or are being considered for publication by journal editors.

There are, in addition, several less tangible, though no less important, results of this project. 1) Experience and training were provided through research assistantships for graduate students, four of whom completed doctorate dissertations on problems in programmed instruction; 2) many instances of exploratory research have prepared the investigators for lines of research that are continuing; 3) dissemination of information by publication, correspondence and direct discussion with the many visitors assisted others in beginning their experimentation with programmed instruction; and 4) the project provided background for a book under preparation by the principal investigator in the technology of education.

# RESPONSE CONTINGENCIES IN TEACHING-MACHINE PROGRAMS[1]

James G. Holland

Programmed instruction rests on the simple, widely accepted principle that when behavior occurs and is followed by a reinforcing event, that behavior is learned. Programming is the arrangement of conditions of *differential* reinforcement such that, at each step, a correct answer will be reached only if the precursory behavior intended by the programmer has occurred. Attaining an answer which the subject considers correct(perhaps, because of confirmation) is said to be contingent upon the response or responses which have had to precede the answer. Anything different from, or less than, the full behavior supposedly taught in a given item or sequence of items must not enable a correct answer. While a public (overt) answer is usually required, the responses on which the answer is contingent are usually private in the form of silent reading, observing graphic material, or solving a problem. Only those aspects of the material that contribute to the subject's ability to answer correctly are programmed; other material unrelated to the answer is not programmed. Therefore, a low error rate does not necessarily indicate an adequate program since the answer may be highly cued in ways unrelated to the major content of the items. However, the second consideration in differential reinforcement does require a low error rate. Differential reinforcement occurs only if the subject is successful in meeting the contingency. Therefore, in a program the precursory behavior must be possible. A low error rate is needed in a program, but it makes considerable difference how that low error rate is obtained.

An experimental demonstration of the importance of these two variables was reported in an unpublished paper (Holland, 1960) in which thirteen sets (377 items)of a prepublication version of the program entitled *The Analysis of Behavior* (Holland & Skinner, 1961) were altered in two different ways. In all cases the content was identical to the normal program except that in each item a different word was chosen to be left blank. In one experimental version, the answer was highly cued and easy to answer, but the cues which determined the answer had as little as possible to do with the material the item was designed to teach. A second experimental version had blanks which were undetermined and quite difficult to answer, thereby providing a high error rate. The words left blank for the subject to complete were ones which would have indicated mastery had the student been able to answer correctly. These two experimental versions were compared with the normal version which reflected the author's attempt to obtain a low error rate, while at the same time preventing subjects from answering correctly for inappropriate reasons. In a fourth condition there

[The JOURNAL OF PROGRAMMED INSTRUCTION, 1964. In Press.]

were no blanks, and the student was asked to read the completed statements presented in the teaching machine. On a post test the group using the normal program made significantly fewer errors than did any of the other three groups. The present paper reports a replication of the earlier experiment.

## Method

Experimental programs. The same thirteen sets of 29 items (total equal 377 items) were used. These were Sets 9-21 in the pre-publication version. Equivalent material in the published program is found in Sets 7-16 covering the introduction to operant conditioning, positive and negative reinforcement, variables affecting speed of conditioning, accidental contingencies and superstitious behavior, shaping, differential reinforcement, and successive approximations. After an item analysis, materials for Conditions B and C were somewhat altered from the previous experiment to make the answer even less relevant to the critical material in Condition B and to make the response less determined (error rate higher) in Condition C. Material for Condition A, the normal program, and Condition D, the version with completed statements, were identical to comparable conditions in the previous experiment. The four programs differed only in the choice of words omitted for the subject to supply as a completion response. Table I presents examples of one item as it appeared in each of the three programs. The total number of blanks within each set of 29 items was constant for Groups A, B, and C, though a given item might differ in the three versions as to the number of blanks.

Subjects. The teaching-machine program was part of the assigned work in a graduate course in educational psychology given at the Harvard Graduate School of Education.[2] The 28 students, registered in the course, automatically took part in the experiment when they reached the experimental sections of the program. After completing the experiment, each subject was advised on the basis of his post test performance which portions of the program he should review to offset inadequacies caused by experimental conditions.

Table I

Sample item as it appeared in each condition

Lesson 11, item 5

Condition A:

Reinforcement which consists of presenting stimuli (e.g. food) is called positive reinforcement. In contrast, reinforcement which consists of terminating stimuli (e.g., painful stimuli) is called ______ reinforcement.

Answer: negative

Condition B:

Reinforcement which consists of presenting stimuli (e.g. food) is called positive reinforcement. In contrast, reinforcement which consists of terminating ______ (e.g., painful stimuli) is called negative reinforcement.

Answer: stimuli

Condition C:

Reinforcement which consists of presenting stimuli (e.g. food) is called 1) ____ reinforcement. In contrast, reinforcement which consists of terminating stimuli (e.g., painful stimuli) is called 2) ____ reinforcement.

Answers: 1) positive 2) negative

Apparatus. The Harvard Disc teaching machine was used in the Harvard Self-Instruction Room (cf. Skinner, 1958). The machine automatically presents one item at a time; the subject answers by writing on an answer strip. Moving a lever exposes the correct answer and simultaneously advances the subject's answer under glass where it cannot be changed. A second movement of the lever presents the next frame. Each item was presented only once; the machine's capacity of returning incorrectly answered items was not used.

Procedure. All teaching-machine work, including the experimental portion, was done in the Self-Instruction Room which had 10 individual cublices(cf. Skinner,1958). Trained attendants issued sets individually, kept records, and monitored the experiments. Subjects used the machines at their own pace and at times of their own choosing. The first eight sets were used in the standard, unaltered form by all students; thereby permitting the subjects to become accustomed to the machine and the program before experimental treatment. For Sets 9-21, the students used one of the four alternative forms depending on previous random assignments to groups. The group N's were slightly uneven (Condition A, 6 S's; Condition B, 8 S's; Condition C, 7 S's; Condition D, 7 S's) because the initial class list had one student who was not registered when final study cards were turned in 10 days after the course began.

After completing Set 21,all subjects took the same 29 item test (Set 22) in the machine. The test covered concepts developed in Sets 1-21. It was labeled a diagnostic test and was designed in part to reveal what review might be necessary. There was no implication that this test or any other teaching-machine records would be used for assigning course grades. In addition,a mimeographed completion test was administered in the classroom. The test consisted of the appropriate portions of a test used in an earlier study (Holland & Porter, 1961). The classroom test was constructed by choosing from the program criterion items classified as to "difficulty" in that they were from one of four different levels of error rates within the program(0-20%, 21-40%, 41-60%, and over 60%). There were too few items of the higher error rate to permit having equal numbers of items in each category. Only four items were available for the highest level and seven items for the second highest level. The lowest and second levels had 17 and 14 items respectively.

Subjects were instructed to record the time on the first and last item of each set. These records provided the measures of time to complete the experimental material. The precision of this measure was felt to be sufficient for its secondary role in this study but was less accurate than would have been possible with automatic recording. The data indicate a tendency to round clock readings to the nearest multiple of five minutes.

After Set 22, the test set, a questionnaire was administered to measure subjects' attitudes to the program they had used.

## Results

Program error rate. The experimental alterations resulted in the expected drastic changes in error rate within the program. Altering the normal version to make answers easy but independent of the critical content reduced the error rate from 14.2 percent to 5.1 percent. Choosing blanks that were highly undetermined increased error rate to 73.3 percent.

Machine administered post test. The results of the earlier study were fully confirmed (see Figure 1). The normal program (Condition A) with answers both dependent on the critical content and well determined resulted in fewer errors than did the other three conditions on the portion of the criterion Set 22,testing on the experimental portions of the program. On these test items,there were 18.3 percent errors for the normal program compared with 39.4 percent for the version of the program with trivial prompting, 40.7 percent for the version of the program with undetermined answers, 38.6 percent for the "reading" control group. The Mann-Whitney U Test permits rejec-

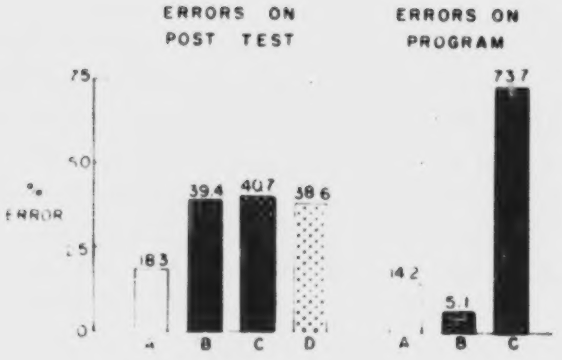


Fig. 1. Percent error on machine administered post test and in the program for four experimental conditions. Condition A used the normal program with response determined and dependent upon critical content; Condition B used the program with responses determined but answers relatively unrelated to critical content; Condition C used the program with responses relatively undetermined but dependent upon critical content; and Condition D used the program with complete statements and with no responses.

tion of the null hypothesis at the .005 level for the difference between Conditions A and B; at the .001 level for the difference between A and C; and at .051 for the difference between A and D. No significant differences were revealed among Conditions B, C, and D.

No group differences were found on the portion of the test set covering the first 8 sets which were used in the same normal form by all groups. Had differences appeared on this portion of the test, the adequacy of the random assignment to groups would have been questionable.

Classroom post test. The classroom test results were consistent with the test set in the teaching machine. In scoring this test, items classified as to different difficulty levels were separately tabulated. As seen in Figure 2, the normal program was superior at all levels of item difficulty. Differences among the other three conditions were not significant at any level of difficulty. The apparent separation at the highest level of difficulty cannot be considered reliable because there were so few items in this category. The suggested separation of conditions at higher levels of difficulty cannot be considered established in this study.

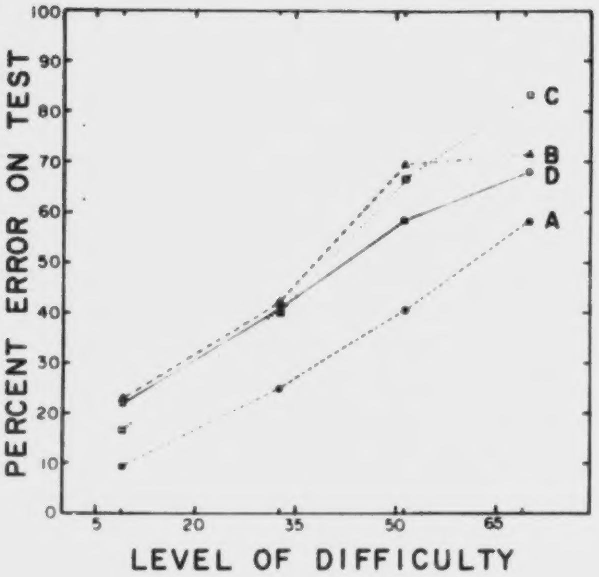


Fig. 2. Percent error on classroom post test as a function of test item difficulty for each of the four conditions. Condition A used the normal program; Condition B used the program with responses determined but answers relatively unrelated to critical content; Condition C used the program with responses relatively undetermined but dependent upon critical content; and Condition D used the program with complete statements and with no responses.

Time. The median times per set showed the shortest time for Condition D (6.3 min.) followed by Condition B (9.9 min.) and Condition A (13.5 min.). Condition C required the longest median time per set (17.4 min.). This rank order of conditions is the same as the rank order of conditions which would result by ranking them in terms of time required to write answers, since answers tend to be the longest for C, the shortest for B, and D has no answers.

Student opinion. The questionnaire results are shown in Table II. The subjects were asked to compare Sets 9-21 in the form they experienced them with Sets 1-8,

Table II

Number of subjects in each experimental condition who gave indicated responses on questionnaire comparing the first eight normal sets with the experimental sets, 9-21.

| | | Condition | | | |
|---|---|---|---|---|---|
| | | A | B | C | D |
| Sets 9-21 were | | | | | |
| much harder | | | | | |
| | -3 | 1 | | 6 | |
| | -2 | 2 | | 1 | 2 |
| | -1 | 3 | 2 | | 1 |
| | 0 | | | | 1 |
| | 1 | | 2 | | 1 |
| | 2 | | 3 | | 1 |
| | 3 | | 1 | | 1 |
| easier | | | | | |
| Sets 9-21 | | | | | |
| taught less | | | | | |
| | -3 | | 1 | 1 | |
| | -2 | | 1 | 3 | 2 |
| | -1 | | 2 | 2 | 3 |
| | 0 | 1 | 2 | | |
| | 1 | 4 | 1 | | 2 |
| | 2 | | 1 | 1 | |
| | 3 | 1 | | | |
| taught more | | | | | |
| Sets 9-21 were | | | | | |
| disliked | | | | | |
| | -3 | | | 5 | 1 |
| | -2 | 1 | 2 | 2 | |
| | -1 | 3 | 2 | | 3 |
| | 0 | 1 | 2 | | 1 |
| | 1 | 1 | 1 | | |
| | 2 | | | | |
| | 3 | | 1 | | 2 |
| liked more | | | | | |

which all subjects used in its normal form. Subjects in Condition A used Sets 9-21 in the normal form which did not differ in programmed rationale from Sets 1-8. The author would judge 1-8 to be slightly easier and not quite so useful as 9-21. This opinion is shared by the subjects in Condition A who found 9-21 a little more difficult and a shade more effective. The subjects in Condition A also slightly preferred Sets 1-8 to Sets 9-21. Since Condition A subjects did not judge 9-21 to be identical to 1-8, judgements by subjects in other groups must be interpreted using Condition A as baseline for comparison. Subjects in Condition B found trivial and highly cued items to be quite easy and to teach somewhat less, though they were no less pleased with the material than were the subjects using the normal program. Subjects using the version with undetermined difficult answers were unequivocal in their collective opinion. They did not like it, found it extremely hard, and thought it taught considerably less. The reading control group (D) was characterized by its wide disparity of opinion, except that it seemed to them on an average to teach a bit less ( especially in view of the baseline judgements of Condition A). While the questionnaire responses are consistent with expectation, it must be remembered that each subject experienced only his own conditions plus the normal Sets 1-8. Had each subject experienced all four conditions and answered the same questionnaire for each, the rank ordering would have probably been sharper.

## Discussion

The results of this experiment and the similar earlier experiment (Holland, 1960) confirmed the principle of program design that behavior engaged in to arrive at the correct answer is more reliably learned. Textual material which does not have to be read or otherwise acted upon to reach the answer cannot be considered programmed (although, of course, some incidental learning occurs just as in ordinary textbooks). Only those things that must be done to answer correctly are a part of the program. Similarly, the material is only programmed when a subject can meet the contingency ( can obtain the correct answer). Items which are like difficult test (Condition C) do occur in published programs, but not so frequently as the "over-prompted" items (Condition B). Neither reflects the proper use of the behavioral technology.

In an ideal program responses would be thoroughly determined(error rate zero)and mastery of every phrase in the program would be necessary for the subject to answer correctly. A program will necessarily teach perfectly if terminal behavior has been precisely specified, and if each item in the sequence (including the criterion items) is both always answered correctly,and only answered correctly after mastering the appropriate material. While this ideal is never completely met, programs need not fall so far short as they often do. To be sure, error rates are usually low, but they are too often low for reasons that circumvent the behavior supposedly being established (as is the program in Condition B). Often this deficiency is obvious; many programs call for the same response term in item after item and others give the first letter,last letter, and number of missing letters of the answer. Moreover, items are often designed so that a consistent phrase prompts the answer. For example,in the section of the program for Condition B concerned with condition reinforcement, each time the word "condition" or "unconditioned"appeared they were followed by the blank and for either the answer was "reinforcer." But in many cases the grammar, syntax or consistent usage in the language can cue the answer unless extraordinary care is taken in the design of a sequence of items. For example,the following item from a generally good mathematics program:

> 230. Another way of saying this is that we are looking for a pair of numbers which satisfy ____ equations at the same time.
>
> Answer: 2 (or both)
> (Lazar, 1962; p. 33)

When this item is presented in isolation it is usually answered correctly, even though the previous items have not been presented. This item is effective as an illustration of the principle because it can be answered without knowing the referent of the pronoun"this." But even if the noun phrase from a previous item was substituted for "this" the item defect would remain; the failure in designing for appropriate response contingency would only be made more obscure.

Even a well written program can be used in a way that permits answering correctly without meeting the intended contingency. This is common when programs are used in a programmed text format which permits the student to peek ahead at the answer and scan other items. Holz & Robinson (1964) found that a programmed text with a vertical format yielded a deceptively low error rate of 4.6 percent as compared with an error rate of 22.2 percent which resulted when the same program was used in a machine which prevented misuse of the program or "cheating."

All other principles of teaching machines and programming as originally stated (Skinner, 1954, 1958) presuppose the design of contingencies for differential reinforcement. Despite this and the general agreement with now platitudinous "the program's the thing," experiments are frequent in which variables, allegedly applicable to teaching machines,are investigated using material which cannot be considered programmed in the sense described above. There have, for example,been a large number of so-called response mode studies which have compared overt responding and either reading completed statements or "thinking"answers without writing. Many of these studies have been criticized in terms of the shortness of the programs used, insensitivity of the tests and inadequacy of the programs ( cf. Cummings & Goldstein,1963;Holland, 1964). The present study has implications for this problem because Condition D, used here as a baseline control condition,corresponds to one of the response mode conditions in many response mode studies. Only the normal program showed a significant advantage for an overt response. The two "non-programmed" conditions (B and C) which required responses and superficially resembled programs were not significantly different from the no-response situation (D) on post test performance. If the response is unrelated to the critical content of an item,it is hardly surprising that responding does not affect the learning of the content. Moreover, it is no more surprising that when the response is undetermined, there is no advantage in overt-responding when the correct response can't be given anyway.

## Summary

A 377 item section of a psychology program was altered by choosing different words or phrases to leave blank for the subject to supply. One version had highly cued answers which were largely unrelated to the major content of the item. A second version had answers which were highly related to critical content but were undetermined, resulting in high error rate. These two experimental modifications resulted in poorer post test performance than the normal program in which error rate was fairly low but correct answers could be reached only by mastering the material in the item. Only the normal program was superior to reading completed statements which required no response.

## REFERENCES

Cummings, Allana Q., & Goldstein, L. S. The effect of overt and covert responding on two kinds of learning tasks. New York: The Center for Programed Instruction, Tech. Report, 1962.

Holland, J. G. Design and use of a teaching-machine program. Paper read at Amer. Psychol. Ass., Chicago, Sept., 1960.

Holland, J. G. Research on programming variables. In R. Glaser and J.H. Reynolds. Teaching machines and programmed learning: II: Data and directions. Washington, D.C.: National Education Association, 1964.

Holland, J. G., & Porter, D. The influence of repetition of incorrectly-answered items in a teaching-machine program. J. exp. Anal. Behav., 1961, 4, 305-307.

Holland, J. G., & Skinner, B. F. The analysis of behavior. New York: McGraw-Hill, 1961.

Holz, W. C., & Robinson, J. A note on machines in a technology of education. J. prog. Inst., in press.

Lazar, N. Verbal problems in Algebra. Chicago: Encyclopedia Britannica, 1962.

Skinner, B. F. The science of learning and the art of teaching. Harvard Educational Review, 1954, 24(2), 86-97.

Skinner, B. F. Teaching machines. Science, 1958, 128, 969-977.

## FOOTNOTES

1. This study was conducted as a part of Title VII of the National Defense Act under the Office of Education, Grant No. 7-31-0370-051, and used the facilities of the Harvard Committee on Programmed Instruction supported by a grant from the Carnegie Corporation.
2. The author wishes to thank Douglas Porter, the instructor of this course, whose cooperation made this study possible.

PRECEDING PAGE BLANK- NOT FILMED

# EVALUATING TEACHING MACHINES AND PROGRAMS

## James G. Holland

MOST PERSONS CONCERNED with education today have heard of teaching machines and are familiar with many of their advantages and disadvantages.[1] The information usually comes either from the all too abundant popular literature or the often too technical scientific literature. Treatment in popular magazines and newspapers is usually more entertaining than informative or accurate, whereas theoretical and experimental work often requires expert interpretation. Technical reports often discuss fundamental principles without making clear how they can best be applied. Furthermore, many of the technical issues have more implications for the future than for the present.

Nevertheless, the current literature suggests that programs for teaching machines provide carefully prepared material, using knowledge about learning gathered from a long history of laboratory research. The program emulates the careful tutor, who analyzes the subject matter and presents problems which progress through the topic in small steps. The material is designed to keep the student an active "learner," and the questions are designed so that the student must master each step to answer them correctly. Further, the feedback to the programmer provided by the student's response enables extremely detailed revision of the learning material. Finally, it is an advantage to have a machine which permits each student to progress at his own rate, moving as far and as fast as he can. Classes no longer need to be geared to the slow or medium student.

In spite of this background, the educator is left with a large number of important questions. Although the potentialities are clear, should the educator immediately try to fill his classes with teaching machines? Should he get programs for all his courses? How can he decide which of several machines to buy? Can he successfully use the program without a machine and thereby avoid the expense? Are there differences in programs, and if so, how shall he evaluate them?

### *Evaluation of Programs*

The quality of programs presently available or under development differs greatly. Claims are made by the zealot that any program will be better than

---

Preparation of this paper and much of the research upon which it is based were supported by the Carnegie Corporation and the Educational Media branch of the Office of Education.

[1] Those interested are urged to examine A. A. Lumsdaine and R. Glaser (Eds.), *Teaching machines and programmed learning: A source book.* Washington, D. C.: National Education Association, 1960. Parts I, III, and V are relevant to the behavioral technology of teaching machines as discussed in the present paper.

[Reprinted from TEACHERS COLLEGE RECORD, Vol. 63, No. 1, October 1961.]

other present day techniques because of the inherent advantages provided by teaching machines (immediate reinforcement, the student moving at his own rate, etc.). But it is quite possible to write a program which is worse than textbook or classroom procedures. There will be an increasing number of programs and machines offered for sale, some quite good, others fair, and still others little more than hasty efforts at exploitation. Indeed, some programmers today seem to have taken seriously the axiom in one of James Thurber's *Fables:* "Don't get it right, just get it written." Hence the dilemma: How is the educator to decide whether or not a program is worthy of adoption?

The first consideration in evaluating a program is whether or not the subject matter embodied in the program is correct and adequate. Many persons who have become familiar with this new technology have proceeded to apply the method without proper attention to the subject matter. A psychologist who has just taken his first course in Russian may begin to try programming Russian. Others, with only the normal college background in mathematics, may begin to program mathematics. Deciding about this aspect of a program is in many ways similar to deciding which of several books to use. One would not adopt a book filled with errors. Neither should one adopt such a program. It is common today to find programs which contain much incorrect material. But of equal importance, the "content," or what the course is teaching, should represent, in all cases, the newest orientation to the curriculum. New substantive developments in the teaching of mathematics, physics, chemistry, languages, etc., should not be sacrificed because of this new technique.

By and large, the best programs will be written by subject-matter experts who have learned how to program. A good programmer, like a good teacher or textbook author, must both be sensitive to the student's abilities and know his field thoroughly. The person who is able to write a satisfactory textbook should, after a short bit of training, be the best prepared person to write an effective program. Since a major criterion for evaluating programs remains the adequacy of the coverage of the subject matter, there is much in common in the evaluation of textbooks and programs. As imponderable as this may be, there is no easy substitute for the reputation of the author, the reputation of the publishing house, and a careful examination of what is being taught.

Quite apart from considerations of subject matter, there are new and important bases for evaluating programs. It is important that the program be well done from a technological standpoint. The ultimate test of programming technique is its empirical success in teaching students. Judgment is still required, however, as to whether a particular program merits the time, effort, and expense of a preliminary use. While we cannot say at the moment that we have all of the answers regarding programming, we *can* specify some general principles which characterize good programs. An adequate program usually has a careful sequence of items arranged in a *gradual progression* of finely graded steps, culminating in items which, standing alone, would be adequate test items for the material in question. The efficacy of such a gradual progression has been demonstrated time and again in the laboratory. Complex behaviors are effectively established in this way, and methods which ignore gradual progression, thereby presenting difficult problems and no adequate preliminary training, are successful only in so far as accidental and uncontrolled contingencies allow some students to reach mastery. With a proper gradual progression, the student may be correct on practically all items. The only unequivocal evaluation of the gradual progression and resulting error rate is to test the program in classroom use. Nevertheless, the experienced teacher may be able to form some preliminary judgment about the adequacy of the gradual progression by simply working through the material himself. In such an evaluation he should look for two things.

First, do the late items in the program, or in a subsection of the program, adequately reflect the behavior the teacher wants established? Second, are the items throughout the program arranged in a sequence which should yield a consistent low rate of errors?

### *Items and Errors*

Not every item designed to produce a low error rate is necessarily effective in teaching. The poorly informed programmer often writes items which can be easily answered without the student's learning what that item is supposed to teach. This misses the point of behavioral control required in good programming. The answer to an item must be one which can be given by the student only after he has gone through the behavior the programmer intends to teach, whether that behavior is simply reading other parts of the item, deducing the answer, inducing a principle, or any of a variety of other tasks. Although the programmer's only contact with the student's behavior is in the written answer, most of the behavior being taught is the covert behavior which occurs in reading the material, reasoning from it to a solution, relating it to other material, etc. Hence, in the completion type of item, the portion left for the student to complete is an extremely important choice point for the programmer. All the various aspects of programming require the student to give the correct answer for the right reason. The response required of the student is often chosen only to assure that the critical content of the item, or perhaps a panel, is carefully read or observed. For example, success in answering the problem may serve to insure the student's careful observation of the material actually in front of him at the moment. A term may be defined in the first part of an item, and an example in the remainder of the item may require a response which can only be given if the definition has been carefully read. Or within a single item, a principle may be illustrated, and the student may demonstrate by his answer a satisfactory adduction of the principle. Other cases involve showing graphs, diagrams, pictures, or descriptions in the form of a panel which is before the student the entire time he works through a lesson. Individual items are then so designed that when the student can provide the correct answer, he has been forced to observe the critical details; therefore, the choice of the word left blank is *not* arbitrary. While the student must be able to answer it correctly, in doing so, he must have "gotten the point" which that particular item attempts to teach.

We have seen that there are two major considerations in considering the response word in the teaching-machine item. One is that of keeping the error level at a minimum. The second, more commonly violated, principle is that of relating the blank to the critical *content* of the item, thereby assuring that the student master the content in order to answer the question. It is rather difficult to write items which simultaneously meet both of these criteria, but this is the stuff of which an adequate program is made.

These two concepts can be illustrated by material from a psychology program prepared by the author. Table 1 contains representative examples of the alternate versions of the material. In all cases, the same basic items were used with different portions left blank. Program A

Table 1

Sample Program Material

| | |
|---|---|
| *Lesson 9, item 2* | |
| Program A:<br>A technical term for "reward" is reinforcement. To "reward" an organism with food is to ______ it with food. | Answer:<br>reinforce |
| Program B:<br>A technical term for "reward" is reinforcement. To "reward" an organism with food is to reinforce it with ______. | Answer:<br>food |
| Program C:<br>A technical term for "reward" is 1) ______. To "reward" an organism with food is to 2) ______ it with food. | Answers:<br>1) reinforcement<br>2) reinforce |
| *Lesson 9, item 7* | |
| Program A:<br>Unlike a stimulus in a reflex, a reinforcing stimulus ______ act as a stimulus to elicit the response. | Answer:<br>does not |
| Program B:<br>Unlike a stimulus in a reflex, a reinforcing stimulus does not act as a stimulus to elicit the ______. | Answer:<br>response |
| Program C:<br>Unlike a stimulus in a reflex, a ______ ______ does not act as a stimulus to elicit the response. | Answer:<br>reinforcing stimulus |
| *Lesson 11, item 3* | |
| Program A:<br>If, instead of presenting food after the pigeon pecks the key, a loud noise is turned *on*, the rate of pecking will *not* increase. Presenting a loud noise ______ a reinforcement. | Answer:<br>is not |
| Program B:<br>If, instead of presenting food after the 1) ______ pecks the 2) ______, a loud noise is turned *on*, the rate of pecking will *not* increase. Presenting a loud noise is not a reinforcement. | Answers:<br>1) pigeon<br>2) key |
| Program C:<br>If, instead of presenting food after the pigeon pecks the key, a loud noise is turned *on*, the rate of pecking will *not* increase. Presenting a loud noise is ______. | Answer:<br>not a reinforcement |

Table 1—*Continued*
Sample Program Material

| | |
|---|---|
| *Lesson 11, item 5* | |
| Program A: | |
| Reinforcement which consists of *presenting* stimuli (e.g. food) is called *positive* reinforcement. In contrast, reinforcement which consists of *terminating* stimuli (e.g., painful stimuli) is called ______ reinforcement. | Answer:<br>negative |
| Program B: | |
| Reinforcement which consists of *presenting* stimuli (e.g. food) is called *positive* reinforcement. In contrast, reinforcement which consists of *terminating* ______ (e.g., painful stimuli) is called *negative* reinforcement. | Answer:<br>stimuli |
| Program C: | |
| Reinforcement which consists of *presenting* stimuli (e.g. food) is called 1) ______ reinforcement. In contrast, reinforcement which consistes of *terminating* stimuli (e.g., painful stimuli) is called 2) ______ reinforcement. | Answers:<br>1) positive<br>2) negative |

was the "correct" program which had been repeatedly used in teaching a behavioral science course at Harvard University. Programs B and C are experimental versions of the very same items with a different word left blank for the student to complete. In the case of program B, the word omitted is an extremely easy one, providing a very low error rate, but, in so far as possible, a word was chosen which was *not* closely related to the critical content of the item. Program C, on the other hand, has highly relevant but very difficult material omitted. Should the student be able to complete the question in program C, he would have all of the critical content of the item. However, the word or words omitted were so chosen that the student was seldom able to answer the item adequately. Experimental use of these alternate versions has demonstrated the superiority of program A. The other two are in fact no better (perhaps worse) than reading the complete statements without having to provide an answer.

It is probably an adequate rule of thumb to say that any portion of an item which is not necessary for the student to arrive at a correct answer cannot safely be assumed to be taught by the item. Although such "extra" material may occasionally be used for stylistic reasons, only material directly relevant to the required answer is truly a part of the controlled or programmed teaching sequence.

## *Learning Abstractions*

An additional principle of programming relevant to evaluation is that of fully developing a concept or, more technically, an abstract response. Good programs do not simply provide for the memorization of principles or other rote learning. Good programming yields a degree of mastery that permits the student to "think" in the ways in which the teacher intended. People often demonstrate an erroneous idea about teaching machines by saying it is all right for rote learning but not for developing real "understanding" or "insight." To establish a concept thoroughly, the program must use as wide a range of examples illustrating the concept as is possible. Variety of examples, not rote drill, is the way to build the student's ability to behave appropriately in the markedly different settings to which the subject matter in question may be applicable. It is this ability which has rather loosely been designated as "understanding" or "insight." The principle of varied examples applies whether we are talking about such concepts or technical terms as reinforcement, force, and electrons, or such simpler concepts as letter patterns in the preliminary stages of learning to read, varied positions and phonetic contexts in training children to hear differences in sounds, and the basic sounds of a foreign language. Similarly, the student who has learned not to say "he don't" may still say "it don't," though both result from the same grammatical error. It is not sufficient to learn the verbalization of a principle and a small number of examples. To master a concept, the student must have a wide range of examples with all of the irrelevant properties varied.

Superficial discussions of programming principles often give the erroneous impression of an over-mechanical type of instruction. The use of programming principles is not an excuse for poor style or dull, humorless, mechanical writing. Some programs are written in this highly mechanical form and therefore should quite correctly be rejected by educators. The principle of varied examples in establishing an abstract response should, when properly adapted, help to avoid the dull and the mechanical. Interesting and aesthetically pleasing material is as relevant to a program as to a textbook.

## *Testing and Revision*

Finally, a program should be tested on a sufficient number of students and the data used in revising the program. A unique advantage of teaching-machine programs is that the author can readily obtain valuable information on the specific deficiencies in his presentation. The students' answers may reveal ambiguity in the items and the means for correcting them; they may point to portions of the program which are moving too rapidly or uncover erroneous assumptions about the students' backgrounds. In addition, the answers may indicate gaps in the program. Incorrect answers may identify places at which the program has moved too rapidly or the portions in which the program has left the student unprepared for the next step. Such difficulties can readily be remedied because the source of difficulty is so easily spotted in the students' responses to each item. Similarly, such feedback may show the advisability of adding a little elementary math to the beginning of a physics program; or in a reading program, an assumption about a child's ability to identify letter patterns may be shown to be erroneous. Furthermore, the preliminary use of the program may suggest the places in which additional prompts are needed or the places in which the programmer should use some new techniques. Although there may be programs in which an author deliberately plans for the student to make errors on certain specific items, the making of *unexpected* errors indicates some deficiency in the program. That deficiency can and should be corrected through revision, preferably done before publication.

This characteristic of programming is distinctive. The usual techniques of control used by teachers and authors provide the information the student is supposed to acquire and leave the technique of acquiring the material entirely to the student. The student is under a threat of a final examination and is simply exposed to a text or lectures with assurance that he must do something in order to learn enough to pass the course.

In contrast, the programmer takes full responsibility to assure that the student comprehends the material as he goes along. Unexpected errors anywhere in the program, including late items which would be equivalent to examination questions, are the fault of the programmer, not the student. This kind of responsibility entails an informed process of revision.

It may be difficult for an educator to determine on his own whether or not a program has been adequately tested and whether proper revision has been undertaken on the basis of such a test. It is hoped that publishers will publish a summary of the results of pretests, indicating the modifications introduced and the extent of the revision. Both ethically and technically, such a practice seems essential.

Unfortunately, in evaluating even such information as may be given, certain ambiguities may arise. Many misleading statements have been made about the number of students necessary in testing a program and about the error level which the program should contain. It is common practice in writing the first version of a program to try it on a few students and then to do some immediate revising. This does not constitute an adequate test even though it is an important step in the production of a first version of the program. As for error level, the program should have been tried on a sufficient number of students, representative of those the program is intended for, to detect the rate of error at which the programmer has aimed. If the author is striving for an error level of 5% or lower on each item, then he must use a sufficient number of subjects to identify such a rate. Testing with thousands of students is unnecessary, expensive, wasteful, and should not be expected, but a sufficient number of students must be used to give a reasonably reliable indication of the error level for each item.

This problem is further compounded by the absence of any unequivocal figure for the ideal error level. In programming psychology, I have attempted to correct the program when error level on any item exceeded 5%. (This figure may be about as low as should be expected for items of this type.) I was only partially successful, however, and the present error level in this program is about 10%. The error level, of course, will certainly vary according to the subject matter or type of program. It may be possible in some cases to reach a zero error level, particularly in some forms of nonverbal material. However, in a program for the training of a fine discrimination—for example, in the training of certain musical skills—it may prove inadvisable to obtain a low error level. Such questions can be answered only through informed judgment and direct research. Dogmatic rules will not apply. Furthermore, error level can be deceptive in another way. In obtaining a low error level, it is quite possible for the programmer to violate the principles of controlled teaching in his attempt to get the student to answer correctly. The problems must still be designed in such a way that the student masters the content to be learned when he answers correctly. In poorly designed items, he may give the correct answers without this being the case. While low error level is usually a necessary condition of good programs, it is not a sufficient condition.

## *The Machine*

Reduced to its barest essentials, a teaching machine is a device which exposes a small problem, provides for an explicit response (usually a written one), and gives immediate confirmation as to the correctness or incorrectness of the response. Beyond a doubt, most of the technology involved in teaching machines is the preparation of material for them. The well prepared program, an example of careful behavioral engineering, is the heart or essence of the entire teaching-machine movement.

Because the program is of such great importance, the suggestion is persistently made by many that the machine itself is unnecessary. Researchers have frequently used flashcards or a printed text with items on one page and answers on another. Proponents of flashcards and the so-called "programmed textbook" argue that since the program is the critical thing, machines are irrelevant. Indeed, if the student carefully wrote his answer each and every time before turning the page to confirm the correct answer, no difference would be expected between programmed text presentation and machine presentation. In the widespread use of programmed materials, however, one cannot expect consistently correct handling of the material. There is a tendency for even the best intended student to glance ahead at an answer as soon as he is even momentarily stumped in providing the answer. Looking ahead has been observed to occur with persons working through a program entirely for their own benefit. These people were without any pressure to continue the program and were not penalized for their errors. This "lazy" use of the program in the programmed-text form is guarded against by the machine, which requires a response to be tried before there is an opportunity to see the correct answer. We have seen that the principles of programming require the student's answer to be under the control of the specific design of the item. When the control over the student's answer is sacrificed, as in the case of the programmed text or the inadequate machine, the careful design of the program is wasted. Sloppy or lazy use of the program, in which answers are not written or the student glances ahead before writing the answer, can only make the program less effective.

In the absence of available machines, the programmed text has some use, if employed on a limited scale, for developing new programs and for experimental purposes. For such practical reasons, many persons have begun using programmed texts. But the trend to continue such use is alarming. Publishers often offer programmed texts primarily because it is convenient to avoid changing their general practices. This is a deplorable development when it occurs without at least parallel publication for one or more machines. True, I would encourage the beginning programmer to use the programmed-text form rather than wait for machines or spend the time in developing machines himself. The programmer can profitably push ahead with programming, realizing the limitations of a programmed text and attempting to make sure that the experimental students use the programmed text properly. But the rationalization that this expedient solution is the ultimately correct procedure must be avoided.

In addition to preventing the lazy use of the program, the machine has other auxiliary advantages. It can be cheat-proof, assuring that the student will not simply copy answers without reading the item at all. Some machines may return incorrectly answered items to the student for a second try. Others may make extra hints available after the student has constructed an initial answer and found an additional cue to be necessary. Some machines have counters so that the student is automatically shown the number of correct responses per lesson.

A decision as to which of several machines to purchase requires the careful weighing of many factors. One overriding consideration is that a machine should absolutely assure that the student, if he writes his answer, does so before he sees the correct answer. There are machines which lack this feature and hence are no better than flashcards or programmed texts. Once this condition has been met, however, it then becomes a matter of weighing many considerations against one another. Cost is an obvious factor (currently, machine costs range from \$20 to \$500). Another is whether there is a separate answer strip or whether the student writes directly on the question material, thus consuming it and rendering it unusable for another student. In addition, machines vary considerably in the ease of operation, the size of the answer space, the convenience of the writing surface, the ease of reading the item, etc. Furthermore, not all manufacturers who offer machines can make delivery in a reasonable time. There has been a regrettable tendency to circulate brochures, complete with pictures, considerably in advance of production. Finally, it is important to find out what programs are available for the machine in question. This may well be the deciding consideration for many buyers. Unless someone plans to write his own programs for the machine, it is essential that published programs be available for the model purchased.

### *Evaluative Research*

It is often asked whether teaching-machine techniques are more effective than conventional ones. In the proper perspective, this is a reasonable question, and it even seems to be at the heart of the issue. Attempts at answering it, however, usually become translated into a conventional experimental-vs.-control-group design which can be extremely misleading. A more valuable research effort is that of discovering how we can most fruitfully provide for well controlled teaching. The pseudo-scientific comparison of "teaching machines" with the "conventional methods" does nothing to further this endeavor. Teaching machines do not constitute a single isolated variable. Consequently, when it is discovered that they are better than the "conventional methods," we do not know why and therefore do not know what must be done to improve the next program. Likewise, there is no one single factor which can be described as the "conventional method." Teachers differ from one another; texts differ from one another, and the same teacher differs from one year to the next. Similarly, programs vary in many ways. When such a control-group experiment is performed, the results can rarely be generalized beyond the highly specific case in which the experiment was done.

The proponents of a control group comparison are victims of a false analogy. The control-group design has served well for many problems in science, but it applies only in cases in which the control and experimental conditions each consist of a specifiable, reproducible condition and in which the experimental condition differs from the control in only a single, isolated variable. This is clearly not the case in comparing teaching-machine programs with conventional classrooms. They differ in many respects, and neither constitutes a generalizable condition. I have rewritten portions of a program in several different forms in order to investigate the variables concerned in programming (cf. Table 1). In comparisons of these alternative forms of the program with a form of the very same items read simply as complete statements with no answer required, the two inadequate forms of the program were as bad, or worse, than merely using the material in an essentially non-machine fashion. The adequate, well-designed program, however, was considerably better than this condition of not having to provide an answer. Aside from the basic information that this study offers regarding techniques of programming, it demonstrates that results will differ according to the adequacy of the particular program. Any attempt to make a general evaluation of teaching machines from any simple comparison of teaching machines with conventional techniques suffers from an inadequate program or an atypically good program. The results are invalid in either case. Similarly, the "conventional method" is at least as unspecifiable as the "teaching-machine method." One cannot, therefore, take at face value studies which proclaim the value of teaching machines on the basis of allegedly "controlled" studies using split classes.

The best way for the educator to proceed in the evaluation of the program is first to judge as well as he can whether the program in question fits his requirements and the principles of good programming, and then to introduce it into the classroom on a small scale. He should proceed cautiously, observing the students working on the material, testing their comprehension and speed, noting their reactions, and making his own judgments as to the merits or demerits of the program for the particular needs of that class. This is similar to the procedures ideally followed in introducing a new teacher. By proceeding gradually, he can decide, after weighing all the factors he has observed, whether to extend the use of that particular program or whether to extend his use of teaching machines gradually into other subjects. In the process of this slow and careful adoption of teaching machines, he can solve some of the new administrative problems which teaching machines bring about. These include such problems as the reapportionment of the effort of teachers and machines, the development of techniques to capitalize on the advantages for the fast student who now has free time for additional learning, and the monitoring of the machines. I strongly urge the educator to proceed to introduce teaching machines as soon as he finds programs worthy of his consideration, but he should move quite gradually in doing so.

### *Summary*

Teaching-machine technology offers many possibilities for the improvement of education, but the success of this new method will depend largely upon the informal judgment of the teacher and school administrator. Careful evaluation of programs is required both as to the accuracy and worth of content and the technical merit of the program. In addition, it is advisable to introduce machines and programs only gradually in order to develop truly informed judgment about the merit of the programs and machines in question and, at the same time, to provide an opportunity for solving the new administrative problems created.

# THE INFLUENCE OF REPETITION OF INCORRECTLY ANSWERED ITEMS IN A TEACHING-MACHINE PROGRAM[1,2]

James G. Holland and Douglas Porter

One feature often considered essential to an ideal teaching machine is a mechanism which requires the student to answer each item correctly at least once. The prototype of write-in teaching machines, developed by Skinner, does provide for repetition of incorrectly answered items; but economic considerations have forced many manufacturers and experimenters to omit this feature. Commercial machines presently fail to return incorrectly answered items to the student, and the cost of a machine is greatly increased when a "memory" is included so that items missed on the first trial are repeated. Instead, current machines present each item only once. The omission of this review feature is often rationalized as inconsequential because teaching-machine programs should produce no errors, and, therefore, repetition of items should be unnecessary for "good" programs. It is self-evident that should no errors be made on a program, reviewing missed items would be superfluous. Equally self-evident, however, is the need for repeating erroneously answered items in an extremely "poor" program which generates many errors. For example, one presentation of paired-associate nonsense syllables (clearly one of the most poorly programmed types of material) would not be sufficient for acquisition of a typical, relatively short, fixed list.

This study evaluates the need for repeating missed items in a program with an error rate thought to be typical of programs for college and high school involving written answers and verbal material. The Psychology Program developed by Holland and Skinner was used. This program has had the benefit of several minor revisions and one major revision aimed at reducing the error rate. Results of the major revision indicate that the error rate has been reduced to a little over 10% of the items (Holland, 1960). The following study uses this revised version.[3]

## PROCEDURE

Fourteen graduate students in an educational psychology course were required to use this program in the Skinner write-in teaching machines as part of their course work. These student subjects were divided into two groups of seven each, matched with respect to their grade average in undergraduate study. One group used the entire program in the usual fashion, repeating missed items at the end of each item set until each item was answered correctly. The second group did not use this review feature but answered each item only once, whether or not it was missed on the first trial. Data on learning outcomes were from 3 tests which were given on appropriate sections of the teaching-machine program and were administered in the classroom. The tests consisted of completion items taken directly out of the program, selected to systematically sample both the subject-matter coverage and level of difficulty of the program. An item error count from previous teaching use of the program provided the basis for choosing test items of different levels of difficulty. Items were chosen at four levels of difficulty: The lowest level had items missed by from 0 to 20% of the students of the previous year; the second level, items missed by 21 to 40% of the students; the third level, 41 to 60%; and the fourth level, items missed by over 60% of the students. Since the program had been designed to keep errors at a minimum, it was not possible to have equal numbers of items at all four levels. Because items having a large percentage of error were so rare, only 18 of the total 196 items were available for use in the most difficult category. The others were: 36 items for the 41-60% level; 68 items for the 21 to 40% error level; and 74 items for the 0 to 20% level.

In addition to the three tests at appropriate points in the program, the same tests were given again, 6 months later, in a final examination. This unannounced retest provides data on the retention of the material and, in addition, information on possible differential forgetting by the two groups.

## RESULTS

The results of the first testing are shown in Fig. 1. The horizontal axis shows the level of difficulty of test items based on the percentage of error which occurred during previous use of the program. The vertical axis shows percentage of error on the hour tests. The two groups of *Ss*, review and nonreview, are plotted separately. At all levels of difficulty, the nonreview group missed more items than did the review group. At the lowest level of difficulty, the difference in the percentage of error is only approximately 4%; at the second level of difficulty, it is about 9%; at the third, about 12% difference; and at the fourth, the

---

[1] This research was reported at the American Psychological Association meeting in 1960 in Chicago.

[2] The Office of Education supported this research financially as a part of Grant No. O. E. 719067.

[3] Since completion of this study, the program (Holland & Skinner, 1961) has undergone still another major revision designed, in part, to further lower the error rate.

[Reprinted from JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR, Vol. 4, No. 4, October 1961.]

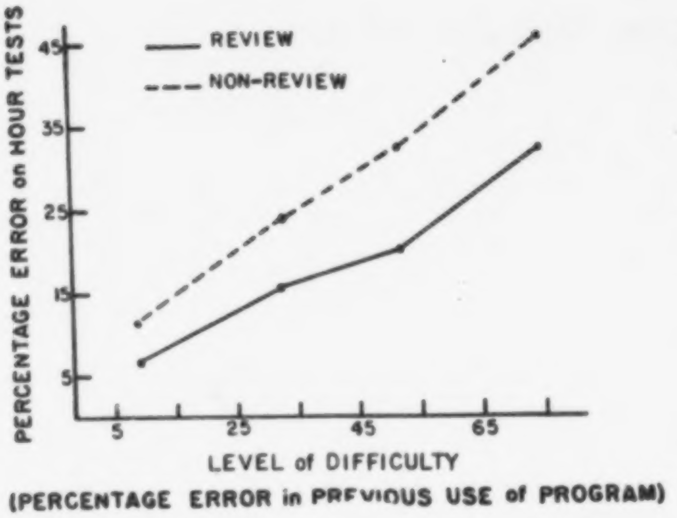


Fig. 1. The average percentage error as a function of difficulty level on the 3 tests administered immediately after the deadlines for completing appropriate sections of the program. Difficulty level is defined as the error rate on these criterion items when presented within the program during a previous use of the material.

difference is about 13%. The difference between these two groups is significant below the 1% level of confidence.[4] In addition, there is a slight trend toward somewhat greater differences between the groups for the higher levels of difficulty. Although this trend is highly suggestive and very reasonable, this interaction effect fails to reach the usual levels of statistical significance.

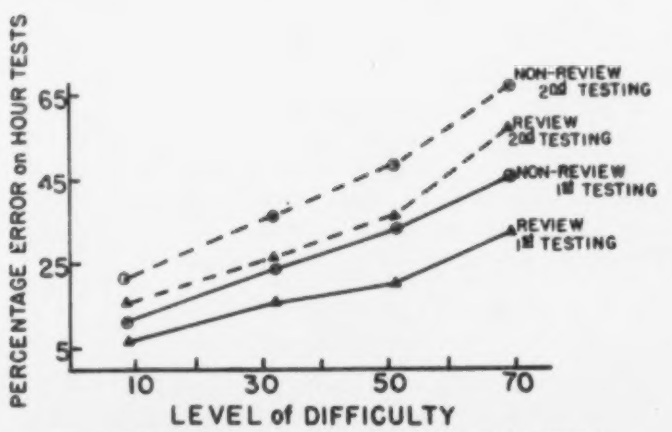


Fig. 2. The average percentage error as a function of difficulty level. The initial test and the retest after a 6-month period are shown separately for each group.

Figure 2 shows results of the retest given 6 months after completion of the teaching-machine work. Again, the level of difficulty is shown on the horizontal axis and percentage of error on the tests on the vertical axis. The two solid lines represent the same data shown in Fig. 1; the two broken lines represent the retest data for the two groups. On retest, the review group is still superior to the nonreview group. Furthermore, both groups have shown some memory decrement at all four levels of difficulty (significant at the .005 level on Wilcoxon matched-pairs test). This memory decrement is better shown by the differences in percentage of errors between the first and second testings. These differences are plotted in Fig. 3. There is a loss of about 10 percentage points at the two lower levels of difficulty; and as difficulty level increases, the differences in percentage error become greater until at the most difficult level there is a loss of between 20 and 25 percentage points from the first to second testing (significant at the .01 level on the Friedman test).

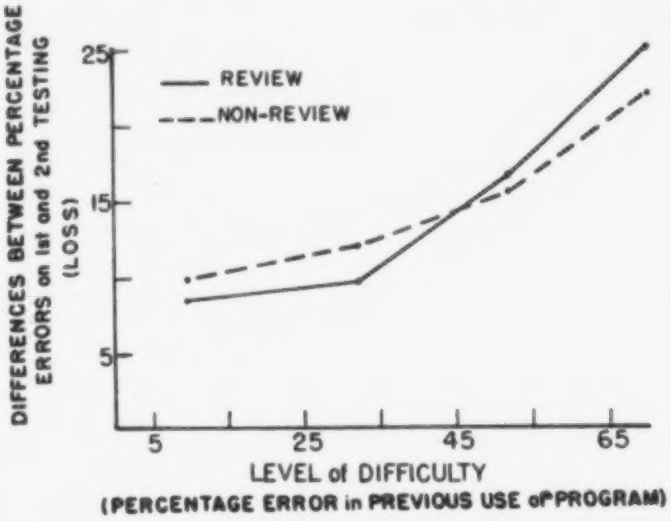


Fig. 3. The average differences in percentage errors between test and retest as a function of item difficulty.

### DISCUSSION

It should be remembered that the levels of difficulty represent the percentage of errors in actual use of the program; the more difficult the teaching items, the greater the memory loss. If the program had been designed so that Ss were ready for each item in its turn and thereby got each one correct, all items would have been at the low difficulty level (0-20% error), and the loss from the first to the second testing would have been limited to 10% or less. These data do not speak well for the "easy-come-easy-go" theory of retention, a hangover from the doctrine of "mental discipline," which has been held by some critics of teaching machines who believe that anything easily learned must be quickly forgotten. On the contrary, the present results indicate that retention is greatest when the program produces a low student error rate.

One final point to notice in the data of Fig. 3 is the almost identical loss from the first to second testings of the review and nonreview groups. Although the nonreview group performs more poorly than the review group on both the initial test and on the retest, there is no difference between the groups in the amount of retention decrement.

---

[4] Wilcoxon matched-pairs test, using as pairs the subscores for the matched subjects on each of the three tests and at each difficulty level.

Although the review condition provides better comprehension of the material, it does require somewhat more time due to the repeating of incorrectly answered items. Table 1 shows the average time per 29-item set for the two groups. The first cycle time is the same for the two groups, but an additional 1.3 minutes per set is needed by the review group to complete the repetition of the erroneously answered items. This results in a total of 1.3 hours over the full 60 sets of items.

Table 1

Average Time in Minutes Per Set of 29 Items

| | Review Group | Nonreview Group |
|---|---|---|
| First Cycle | 12.3 | 12.3 |
| Total Cycle | 13.6 | — |

These results came as a surprise to us. We had expected very little difference between review and nonreview conditions because the error level in this program is relatively low. Furthermore, in reviewing the few missed items in each set, students often stated that they remembered the answer provided without having to read the item on its return, so that any advantage provided by the review feature would be canceled out. Apparently, however, either the memory for missed items is not so prevalent as we expected on the basis of student reports, or else short-term rote memorization of an answer is insufficient to ensure that it will be emitted again under appropriate stimulus conditions as provided by the original teaching-machine item. In the near future, at least, economics will probably prevail, and we can but regret loss of the review feature in commercial machines becoming available. The question then becomes: What can one do to overcome this deficiency; or, how can one use the additional 1.3 minutes required for review to advantage for a nonreview group? One possibility is to use diagnostic tests which will indicate review of whole lessons in areas where the student is weak, or, for a more ambitious programmer, to use parallel blocks of programmed material. A second recommendation is that of Skinner, who suggested a double stage of confirmation for each item. After the student writes his answer, the machine would reveal an additional hint for use if needed; then, if necessary, the student would write a new answer and the machine would provide the final, complete answer. Such a double-stage confirmation might make it possible to ensure that the student answers the item correctly 100% of the time.

Despite all such techniques to compensate for the elimination of the review feature, the best solution is better programs. If a program were so well written that errors seldom occurred, not only would retention be great but the review feature would be superfluous.

## REFERENCES

Holland, J. G. Teaching machines: an application of principles from the laboratory. *J. exp. Anal. Behav.*, 1960, **3**, 275-287.

Holland, J. G., and Skinner, B. F. *The analysis of behavior: a program for self-instruction.* New York: McGraw Hill Co., 1961.

*Received February 9, 1961.*

PRECEDING PAGE BLANK-NOT FILMED

# A NOTE ON MACHINES IN A TECHNOLOGY OF EDUCATION

**W. C. Holz and John Robinson**

Teaching machines evolved from efforts to develop a technology of education based upon the principles of experimental psychology. Such a technology attempts to determine maximally effective teaching procedures empirically, and is guided by the general principles of learning established in the laboratory. In order to apply the principles appropriately as well as to assess the effectiveness of techniques, the learning situation must be controlled. When the material presented is specifiable and the subject's response is unambiguous, the procedures may be modified until satisfactory learning takes place.

A proposition that "learning" cannot occur without a machine, of course, is indefensible. Nor does it seem likely that the technology can do more than effectively harness learning processes which have always existed. The purpose of a teaching machine is simply to provide the necessary control over a sequence of material. The recent controversy over the utility of machines has, however, tended to cloud the issue. Studies which have compared machine and text presentations of programs (cf. Goldstein & Gotkin,1962) emphasize the fact that subsequent mastery tests show no difference between the two methods of presentation. The suggestion is implicit that if there are no differences in the mastery tests, then, in general, machines and texts are interchangeable. The following study shows that other differences do exist between machine and text presentations and further suggests that the differences are related to the degree of control over the learning process. The point is that teaching machines are important, not because they introduce some magic into the teaching process, but simply because they afford control which is not otherwise obtainable.

A 350 item geography program (Haring & Haring, 1962) was administered to 6th grade students who were not otherwise receiving instruction on this topic. All of these students demonstrated lack of prior knowledge of the material by scoring below 40% on a pretest. This test of 36 items was based upon critical content of the program and was readministered as a posttest. The program was presented to 8 students in a Koncept O Graph teaching machine, and to 12 students in a programmed text which used a standard vertical format. The students using the text were equated with those using the machine on the basis of pretest scores. The text was equipped with a sliding shield with which to cover succeeding answers on the page (but not previous items or their answers). Administration was conducted under supervision to groups of 4 when the machine was used and groups of 6 when the text was used.

[Reprinted from The JOURNAL OF PROGRAMED INSTRUCTION, Vol. II, No. 3, 1963.]

Table I

| | Machine | Text |
|---|---|---|
| Errors on Program (Median) | 22.2% | 4.6% |
| Semi Interquartile Range $\left(\frac{0_3 - 0_1}{2}\right)$ | 7.5% | 2.1% |
| Posttest Scores (Mean) | 23.9 | 25.7 |
| Standard Deviation | 5.3 | 4.7 |
| Product Moment Correlation Between Posttest Errors and Errors on Review Set | .61 | .17 |

The first item in Table I shows the error rates on the items. The error rate for the machine presentation was considerably above that found with the text version (median test significant above the .01 level). In spite of the lower error rate with the text presentation, the posttest performances (item 2 in Table I) showed no significant differences ($p > .20$). The third item in Table I shows the correlations between posttest errors and errors on the program's review set, which closely paralleled the test.

Although the students using the text proceeded through the program with few errors, their posttest performances were far from errorless (about 30%). The error rate for the students using machines reflected more nearly the level of performance on the posttest and correlated more highly with it. These results suggest that the program error rate is spuriously low for the text. The explanation seems to lie in the lack of control over the material provided by the text. Only the machine prevented the student from referring back to previous items and their answers. With only slight movement of the shield, the student using the text could expose the correct answer before writing his own. In any event, the error rates were quite different, and in view of this, it seems advisable to consider the form of presentation whenever error rates are used to evaluate or review programs.

It may be argued that although teaching machines are useful in program revision, once an adequate sequence of material is developed a programmed text will be as effective. The resolution of such an argument must await further research with programs of demonstrated teaching effectiveness which cover large amounts of material. If programmed texts are used the way they are intended, then no important difference between text and machine need be expected. But the absence of specific requirements to use the text properly, allows the user to become slipshod and eventually to pervert their use. Volunteer subjects may be cooperative enough to use the text as instructed for an hour's experiment; but this says nothing about the student who uses programmed material day in and day out. The ultimate value of the teaching machine may be the requirement that the student study in exactly the manner prescribed by the programmer. It would be naive indeed to assume that simply because a "machine" is used that all artifacts automatically would be avoided. The so-called "cheat-proof" machine is probably yet to be built and careful supervision can only hope to minimize the problem. But the development of convenient, effective machines is as important a problem for the technology of education as is the development of effective teaching sequences.

## REFERENCES

Goldstein, L. S. and Gotkin, L. G. A review of research: teaching machine vs. programed textbooks as presentation modes. J. prog. Inst., 1962, 1, 29-36.

Haring, L. L. and Haring, Darlene. Latitude and Longitude. Chicago: Coronet Instructional Films, 1962.

# VARIABLY BLURRED PROMPTING IN THE ANALYSIS OF PAIRED-ASSOCIATE LEARNING (ABSTRACT)[1]

**Matthew Israel**

A single verbal response may have several sources of strength. When one or more sources are brought into play, but prove inadequate to evoke a desired response, a supplementary source may be added. If sources combine additively, the amount of supplementary source required should be an indirect measure of the combined contribution to response probability made by the other sources.

This possibility was examined in three paired-associate experiments. S was presented with the German member of a German-English pair. If he could not recall the English, he clarified a blurred image of it until the image prompted a correct recall. Clarification was effected by moving a screen at a slow constant speed towards the plane in which the image was focussed. (See Fig. 1 & 2).

The amount of clarification or "prompt" required for a correct recall is operationally equivalent to a "recognition threshold." The terms will be used interchangeably.

## EXPERIMENT I

Consider a pair whose English member is not recalled until the fourth trial. Does the amount of prompt required on the second and third trials indicate the German member makes a partial contribution to probability of recall? If so, is the contribution greater on the third than on the second?

Each presentation of the list of pairs is termed a "cycle." Experiment 1 showed that thresholds did drop from cycle to cycle (Fig.

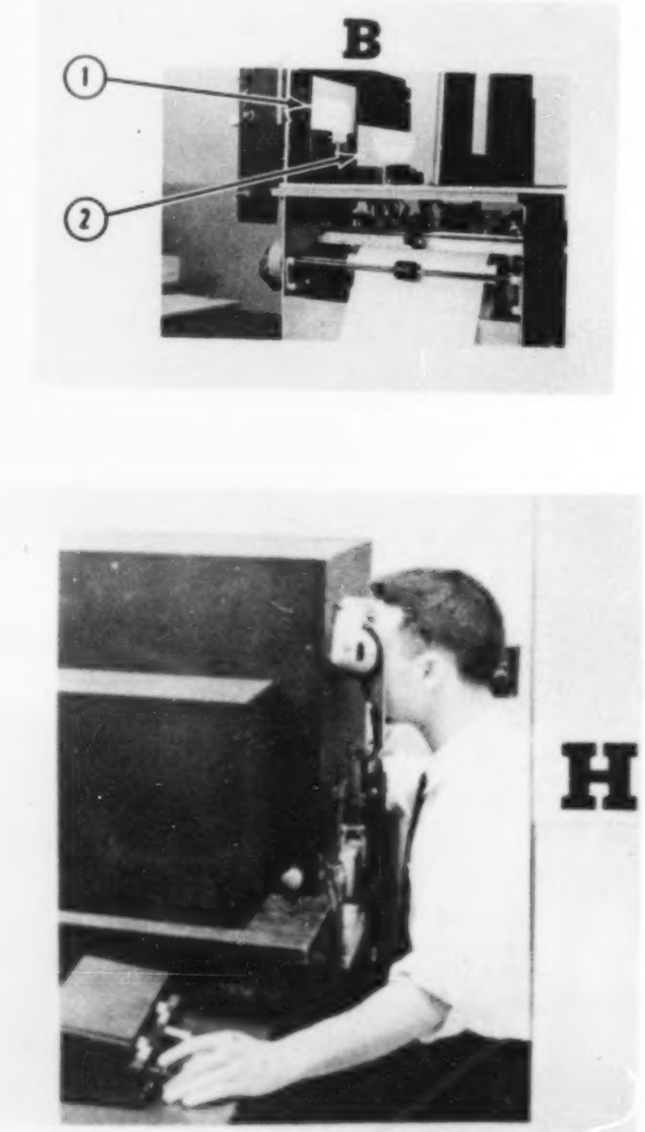


Fig. 1. A. Front side view of prompter, showing fixed screen (1) and movable screen (2).
B. View of Subject operating prompter.

---

[1] The first of the three experiments was published in the Journal of Psychology, 1960, 50, 43-52, under the title Variably blurred prompting: I. Methodology and application to the analysis of paired - associate learning.

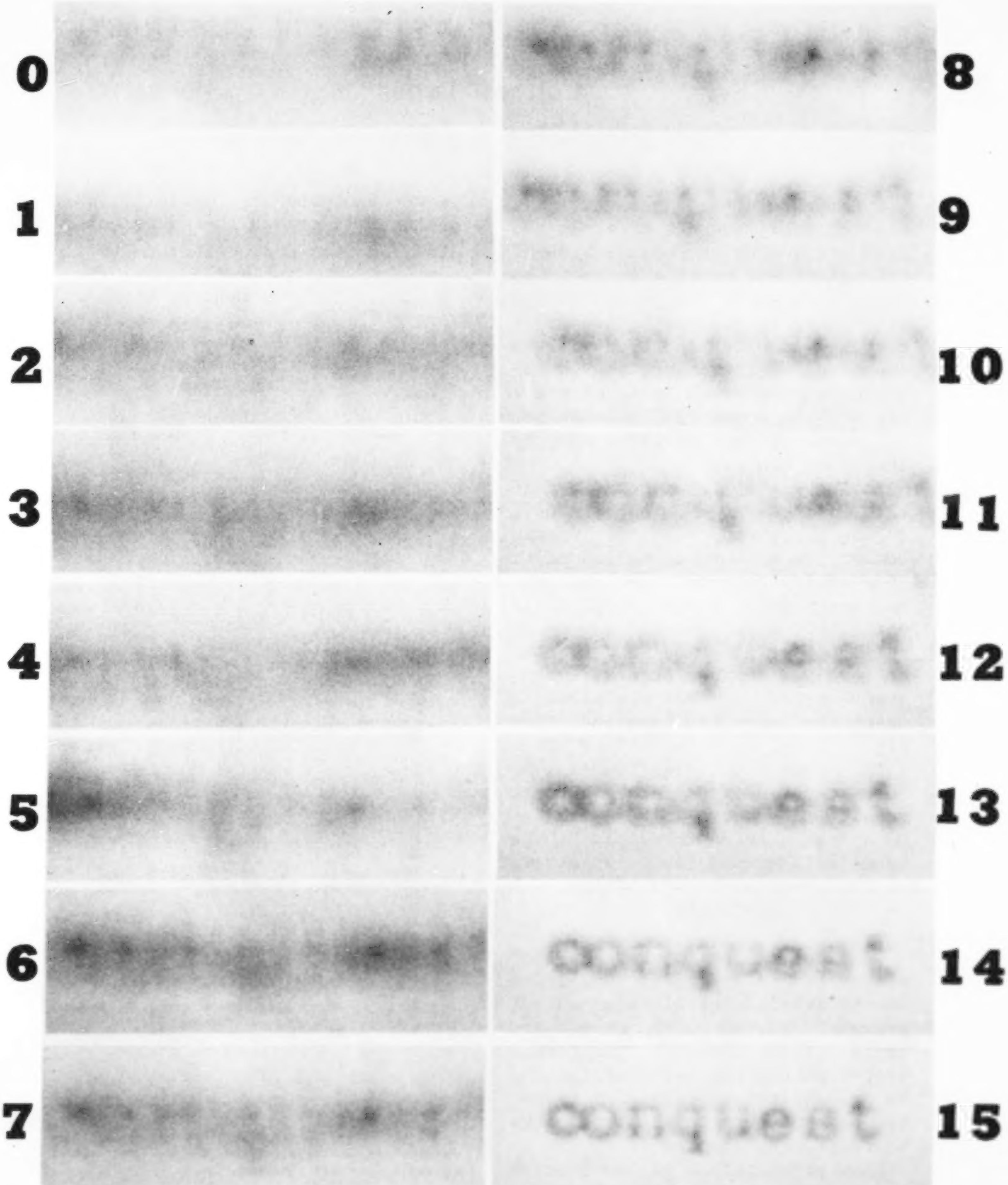


Fig. 2. The word "conquest" at 16 equally spaced stages of increasing clarification. The number beside each blur pattern gives the magnitude of clarification in cm. of screen displacement. Minimum clarification is 0.0 cm. Maximum is 15.7 cm. (not shown). The blur patterns are reproduced in their actual size. By holding the figure about 38 cm, (15 in.) from the eye, the reader may view the blurs approximately at the visual angle they subtended for the Ss of these experiments.

3). However, this drop cannot be attributed entirely to the contribution made by the German members. The English members may merely have increased in probability (apart from coming under the intraverbal control of the German members), causing thresholds to drop; and the threshold measure may have been affected by some blur-discrimination learning. The combined effect of such factors was evaluated by interleaving control trials into the basic set of German-English experimental trials. On control trials thresholds were measured for English words alone.

Control thresholds also showed a cycle-to-cycle drop, though consistently smaller than the experimental. The difference between these drops (the "drop difference") is the best estimate of the partial contribution of the German words.

The drop-difference measure was applied to experimental pairs that could be followed through at least three cycles prior to recall. No clear evidence was found for a cycle-to-cycle increase in the partial contribution.

## EXPERIMENT II

After each prompted recall in Exp. 1, S studied the pair as long as he wished. No comparable studying occurred on control trials. Although the major effect of studying may be to give the German words intraverbal control over the English, studying may also merely make response-members more probable. If so, perhaps the experimental drops in Exp. 1 exceeded the control not because of partial contributions, but rather because the English members of the experimental pairs were benefiting from this added source of response-member probability.

To handle this problem, Exp. 2 introduced a "dropout" control, in which the English word to be recognized was a member of an experimental pair presented on a previous cycle. Drop differences were found in Exp. 2, confirming the results of Exp. 1.

Experiment 2 also included control trials of the kind used in Exp. 1 ("repeating" controls). The drop of the dropout controls was only slightly larger than that of the repeating controls, and of small (.07) statistical significance (Fig. 3). Evidently, little response-member learning is caused by the studying of a pair. Apparently, most response-member learning is generated simply by the presentation of the pairs to S for reading.

## EXPERIMENT III

Experiment 3 attempted to measure the partial contribution of the German words against as level as possible a control baseline. English words were (Exp. 3A) presented alone for recognition for six or seven cycles, then (3B) paired with German words in a paired-associate prompting experiment for four cycles, and finally (3C) presented alone once again for four or five

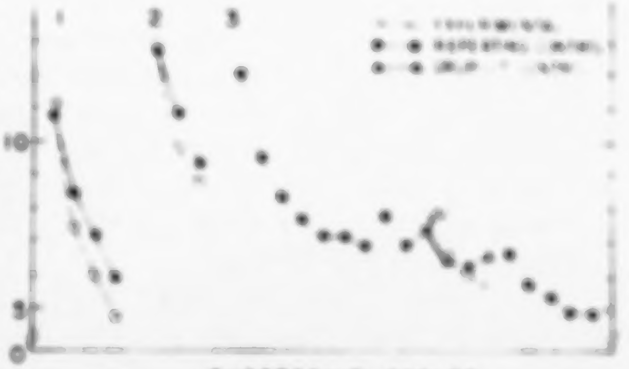


Fig. 3. Pooled curves for Exp. 1, 2, and 3. The numerals indicate the number of the experiment. Mean threshold curves are given. The curves for Exp. 1 and 2 represent, in each case, the pooled data for all Ss -- not merely the Ss common to the two experiments. The dropout control data are omitted from the curves for Exp. 2. Cycle 3 in Exp. 1 and cycle 4 in Exp. 2 are not plotted, since the experimental means for those cycles were based on too few thresholds (16 and 12, respectively).

cycles.

In the latter part of Exp. 3A, threrholds leveled off (Fig. 3). In Exp. 3B, thresholds rose when the German words were introduced. Perhaps the German words had an anxiety effect that included a lowering of the probability of all verbal responses. Or perhaps the textual stimulus for the German word increased its probability, thereby decreasing the probability of all other words -- including the English words.

Threshold differences were not found in Exp. 3B (Fig. 3). Three points may be noted:

(a) When the German words were introduced, thresholds had already dropped to 60 percent of their first-cycle values. This left a smaller, and possibly less sensitive, part of the clarification continuum along which the intraverbal contribution could be measured.

(b) Since Exp. 3A had made the English members quite probable in S's behavior, the range over which the German words could generate a further increase in probability during 3B was much smaller than it had been in Exp. 1 and 2.

(c) Since thresholds rose on the cycle on which German words were introduced, and since that was the cycle from which all drops were measured, the control drops may have been measured from artificially high initial values.

Experiment 3C showed that the threshold rise caused by the German words was not an artifact of the particular cycle on which they had been introduced. It also showed that the threshold leveling at the end of 3A was only temporary. In 3C, thresholds dropped below the levels they had reached in 3A and 3B. Perhaps the leveling at the end of 3A was due to fatigue.

## Discussion

1. The demonstration that partial contributions exist on trials prior to recall contradicts a position recently taken by Rock. His results can be explained without denying partial contributions; furthermore, his measure of learning -- recalls vs. errors -- is not as sensitive a measure of partial contributions as variable prompting.

2. The usual analysis of paired-associate learning in terms of the formation of "associations"--"partial" or "complete" -- is rejected in favor of a functional analysis that takes probability of recall as a dependent variable and investigates the operations which raise and lower it and shifts its stimulus control.

3. The major weakness of prompting is the possible intrusion of unprompted ("time-generated") recalls into the threshold measure. One solution is to require S to wait for a period of time before allowing him to use the prompter. Another is to redefine "prompt" to include the passage of time.

4. Does a prompt (a) increase the probability of the prompted response, (b) decrease the probability of alternative responses, or (c) do both? Is there a distinction? Suppose S is unaware he is being prompted. Is this a case where a prompt can raise the probability of a response without having an elimination-of-alternatives effect?

5. Further applications of prompting include these. (a) In verbal learning studies, a measure combining partial contributions with recalls is possible. (In the present experiments, this could have been obtained by allowing unprompted recalls to enter the ex-

perimental means as "prompts" of zero degree.) (b) In the field of teaching machines, a variable prompter may prove useful for at least two reasons. By putting partial contributions to "work" in prompted recalls, it may speed up the rate at which unprompted recalls can be made. Also, it may have important motivational properties.

PRECEDING PAGE BLANK- NOT FILMED

# THE INCORPORATION OF A PSSC FILM INTO A PROGRAM[1]

**Philip G. Schrag**
**James G. Holland**

The present experiment demonstrates the advantage of applying the techniques of programmed instruction to concepts which need temporal stimulus dimensions best provided by the motion picture medium. In designing teaching machine programs, the stimulus material used is dictated by the nature of the behavior to be established. While ordinary printed statements are common, as they generally are in educational material, the particular skill being established may require other stimulus materials like models, instruments, or laboratory materials to be manipulated. When dealing with time and motion concepts, actual or filmed demonstrations are especially indicated. But the mere passive demonstration by the instructor or by the film does not in itself establish in each student the concept in question. Only if the material is so designed that each student successfully responds in a fashion that assures mastery of each concept or step leading to the concept does the material assure that each student learns. The material then meets the requirement of a teaching-machine program.

It is ironic that vast sums have been spent on curriculum design projects which have limited their pedagogical innovation in the use of filmed presentation. Such presentations have a disadvantage shared by lectures in their failure to directly manipulate the learning process. As a passive device, film and television do not require student mastery of each point before presenting more advanced material. The student during films may be inattentive much of the time, leaving him unprepared to move ahead. In a teaching-machine program, however, attention is assured; frames advance only after they have been answered, and the frames are so designed that they are answered only if the frame's point has been learned.

Gropper (1963) has demonstrated that many of the principles of programmed instruction can be applied to filmed material by asking appropriate questions of the audience and allowing time for them to answer before confirmation. The present demonstration reveals what such techniques can add when applied to an existing excellent educational film which in its original form lacked the features of programmed instruction (the PSSC physics film entitled "Frames of Reference.") Moreover, the present demonstration applies programmed instruction more completely by incorporating the filmed material into a student paced, cheat-proof teaching machine.

## METHOD

Apparatus. A Bell and Howell 16 mm. motion picture projector

---

[1] A portion of this work was supported by a grant from the Carnegie Corporation.

was modifed to include: auxiliary remote light, sound motor controls, and two micro-switches which made contact with either side of the film as it ran through the machine. The film was coded by notches along its edges. When instructed by the printed portions of the program, the student pressed a button which turned on the motor, light, and sound. At the appropriate point, the light and sound would be extinguished automatically (signaling the student to advance to the next printed frame.) The motor would advance the film to the next sequence to be viewed, and then would stop until the button was again pressed.

A TMI MinMax II with Answer-Mate was used to present the printed portion of the program and to provide the student's answer space.

Program. In general, only the portions of the film showing demonstrations are used in the program. The lengthy lecture in the second half of the film and most of the explanations and interpretations in the sound track have been replaced by conventional printed teaching-machine frames. The student begins with the printed program and at eighteen points within the program is instructed to press the start button for the projector to present a demonstration. Undesired portions of the film were automatically bypassed by operating the projector with the bulb off. In the integration of the film and the printed items, the film played a variety of roles:

a) Providing stimulus material for technical terms:

> Now view film sequence A.
>
> We usually think of one thing as being fixed and other things as being moving, but really, all motion is ________.
>
> Relative

b) Enforcing observing the movie, by asking for detailed observation:

> When the cart was standing still, the ball (did/did not) bend right at the bottom of the vertical reference stick.
>
> Did

c) Using the film sequence to confirm his response:

> The next film sequence is going to show how the ball's path looks when viewed from a frame of reference moving WITH THE CART. How would you predict the path will appear from that frame of reference?
>
> A straight vertical line (or equivalent answer.)
>
> Now view sequence D.
>
> Watching the motion from a fixed frame of reference there should be no "unusual" effects. If we ignore the table and the professors, we should see the path of the puck as a * * *.
>
> Straight line.
>
> Now watch sequence H.

d) Asking the student to compare two stimuli he has seen:

> Viewing the motion from a frame of reference moving with the cart, the motion of the ball appeared just the same as in which previous demonstration?
>
> Dropping the ball from the cart when it was standing still.

In addition, between movie sequences standard programming devices were employed to teach concepts.

The Test. A test of terminal behavior was written to test for everything covered by the original film and its sound track. Open-ended questions were used rather than "fill-ins" in order to avoid giving unfair advantages to the students who had received programmed instruction. Before taking the post - test, subjects were instructed: "answer in complete sentences, and try to make your answers as complete as possible."

Subjects. Obtaining subjects with an appropriate background was difficult. The ideal subject would be one who had completed the PSSC course up to but not including the lesson on frames of reference. As a compromise, students were used who were familiar with the basic vocabulary of dynamics, but who had never studied frames of reference or relativity theory. Students were given brief interviews before being accepted. No student who could not define "force" (mass times acceleration) and "acceleration" (change in velocity with respect to time) was accepted, since they would not be familiar with the concepts necessary for the program. In general, accepted subjects were those who had taken one year of physics, either high school or college. (But this varied. Students who had taken Physics I at Harvard were usually accepted, but those who had taken the more intensive Physics 12 were rejected.) Fifteen were given the program; and six were control subjects. All subjects were paid a flat rate of $3.00.

Procedure. To make sure that subjects who had been accepted were familiar with the term "acceleration" as it is used in the film, both program and control subjects were given one-half page of "prerequisite information" to read over. They were told, "If anything here is confusing in any way, ask now and it will be explained to you." No one had any questions.

Control subjects viewed the complete PSSC film. Experimental subjects used the program. The film alone as shown to the control subjects took 45 minutes; whereas the program took between one hour and an hour and three quarters to complete with an average time of an hour and a quarter. Times are, however, deceptive, because of the delay between each frame in operating the particular model teaching machine used here and because of occasional delays while experimental subjects waited for portions of the film to be by-passed.

Most subjects went through the program,as described above, seeing segments of the film. A few subjects, however, were used as controls; they saw the entire film (including an additional twenty minutes excluded from the program), and received no programmed instruction. Both groups took the same post-test.

## RESULTS AND DISCUSSION

The results are summarized in Table I. The median post - test score for subjects using the program was 9.5 out of a possible 10; in contrast, the median score for subjects who viewed the original

TABLE I

Medians and ranges of correct answers on post-test for the subjects using the program and for subjects viewing the film.

| | PROGRAM | FILM ONLY |
|---|---|---|
| Median | 9.5 | 6.25 |
| Range | 6.5 - 10 | 5 - 9 |
| N | 15 | 6 |

PSSC film was 6.25, although one of the subjects in this group scored 9 points. Test performance indicates the program was highly effective.

The results should not be taken as a proof that "programmed instruction" is better than "educational films." The control was used here to provide a base line for evaluation of this specific program. The efficacy of the first version of the program seems well established as compared with the highly developed PSSC film. It is not possible for a simple program or a single film to serve as representative of its class of material. The results do demonstrate that it is feasible to incorporate an existing film into a program when temporal properties are important. This can be done with an existing benefit even when the standard film is an excellent one as in the case of the PSSC film, "Frames of Reference." The benefits of doing so might reasonably be expected to be even greater if the programmer also played a role in preparing the script for the film.

## REFERENCES

Gropper, George L. Why is a picture worth a thousand words? AV Communication Review., 1963, 11, 75-95.

# BASIC SKILLS: Rationale and Procedure

# WHY WE NEED TEACHING MACHINES

## B. F. Skinner

CURRENT suggestions for improving education are familiar to everyone. We need more and better schools and colleges. We must pay salaries which will attract and hold good teachers. We should group students according to ability. We must bring textbooks and other materials up to date, particularly in science and mathematics. And so on. It is significant that all this can be done without knowing much about teaching or learning. Those who are most actively concerned with improving education seldom discuss what is happening when a student reads a book, writes a paper, listens to a lecture, or solves a problem, and their proposals are only indirectly designed to make these activities more productive. In short, there is a general neglect of education method. (Television is no exception, for it is only a way of amplifying and extending *old* methods, together with their shortcomings.)

It is true that the psychology of learning has so far not been very helpful in education. Its learning curves and its theories of learning have not yielded greatly improved classroom practices. But it is too early to conclude that nothing useful is to be learned about the behavior of teacher and student. No enterprise can improve itself very effectively without examining its basic processes. Fortunately, recent advances in the experimental analysis of behavior suggest that a true technology of education is feasible. Improved techniques are available to carry out the two basic assignments of education: constructing extensive repertoires of verbal and nonverbal behavior and generating that high probability of action which is said to show interest, enthusiasm, or a strong "desire to learn."

The processes clarified by an experimental analysis of behavior have, of course, always played a part in education, but they have been used with little understanding of their effects, wanted or unwanted. Whether by intention or necessity, teachers have been less given to teaching than to holding students responsible for learning. Methods are still basically aversive. The student looks, listens, and answers questions (and, incidentally, sometimes learns) as a gesture of avoidance or escape. A good teacher can cite exceptions, but it is a mistake to call them typical. The birch rod and cane are gone, but their place has been taken by equally effective punishments (criticism, possibly ridicule, failure) used in the same way: the student must learn, or else!

By-products of aversive control in education range from truancy, early drop-outs, and school-vandalism to inattention, "mental fatigue," forgetting, and apathy. It does not take a scientific analysis to trace these to their sources in educational practice. But more acceptable techniques have been hard to find. Erasmus tells of an English gentleman who tried to teach his son Greek and Latin without punishment. He taught the boy to use a bow and arrow and set up targets in the shape of Greek and Latin letters, rewarding each hit with a cherry. He also fed the boy letters cut from delicious biscuits. As a result, we may assume that the boy salivated slightly upon seeing a Greek or

[Reprinted from HARVARD EDUCATIONAL REVIEW, Vol. 31, No. 4, 1961.]

Latin text and that he was probably a better archer; but any effect on his knowledge of Greek and Latin is doubtful.

Current efforts to use rewards in education show the same indirection. Texts garnished with pictures in four colors, exciting episodes in a scientific film, interesting classroom activities—these will make a school interesting and even attractive (just as the boy probably liked his study of Greek and Latin), but to generate specific forms of behavior these things must be related to the student's behavior in special ways. Only then will they be truly rewarding or, technically speaking, "reinforcing."

We make a reinforcing event contingent on behavior when, for example, we design a piece of equipment in which a hungry rat or monkey or chimpanzee may press a lever and immediately obtain a bit of food. Such a piece of equipment gives us a powerful control over behavior. By scheduling reinforcements, we may maintain the behavior of pressing the lever in any given strength for long periods of time. By reinforcing special kinds of responses to the lever—for example, very light or very heavy presses or those made with one hand or the other—we "shape" different forms or topographies of behavior. By reinforcing only when particular stimuli or classes of stimuli are present, we bring the behavior under the control of the environment. All these processes have been thoroughly investigated, and they have already yielded standard laboratory practices in manipulating complex forms of behavior for experimental purposes. They are obviously appropriate to educational design.

In approaching the problem of the educator we may begin by surveying available reinforcers. What positive reasons can we give the student for studying? We can point to the ultimate advantages of an education—to the ways of life which are open only to educated men—and the student himself may cite these to explain why he wants an education, but ultimate advantages are not contingent on behavior in ways which generate action. Many a student can testify to the result. No matter how much he may *want* to become a doctor or an engineer, say, he cannot force himself to read and remember the page of text in front of him at the moment. All notions of ultimate utility (as, for example, in economics) suffer from the same shortcoming: they do not specify effective contingencies of reinforcement.

The gap between behavior and a distant consequence is sometimes bridged by a series of "conditioned reinforcers." In the laboratory experiment just described a delay of even a fraction of a second between the response to the lever and the appearance of food may reduce the effectiveness of the food by a measurable amount. It is standard practice to let the movement of a lever produce some visual stimulus, such as a change in the illumination in the apparatus, which is then followed by food. In this way the change in illumination becomes a conditioned reinforcer which can be made immediately contingent on the response. The marks, grades, and diplomas of education are conditioned reinforcers designed to bring ultimate consequences closer to the behavior reinforced. Like prizes and medals, they represent the approval of teachers, parents, and others, and they show competitive superiority, but they are mainly effective because they signalize progress through the system—toward some ultimate advantage of, or at least freedom from, education. To this extent they bridge the gap between behavior and its remote consequences; but they are still not contingent on behavior in a very effective way.

Progressive education tried to replace the birch rod, and at the same time avoid the artificiality of grades and prizes, by bringing the reinforcers of everyday life into the schools. Such natural contingencies have a kind of guaranteed effectiveness. But a school is only a small part of the student's world, and no matter how real it may seem, it cannot provide natural reinforcing consequences for all the kinds of behavior which education is to set up. The goals of progressive education were shifted to conform to this limitation, and many worthwhile assignments were simply abandoned.

Fortunately, we can solve the problem of education without discovering or inventing additional reinforcers. We merely need to make better use of those we have. Human behavior is distinguished by the fact that it is affected by small consequences. Describing something with the right word is often reinforcing. So is the clarification of a temporary puzzlement, or the solution of a complex problem, or simply the opportunity to move forward after completing one stage of an activity. We need not stop to explain *why* these things are reinforcing. It is enough that, when properly contingent upon behavior, they provide the control we need for successful educational design. Proper contingencies of reinforcement, however, are not always easily arranged. A modern laboratory for the study of behavior contains elaborate equipment designed to control the environment of individual organisms during many hours or days of continuous study. The required conditions and changes in conditions cannot be arranged by hand, not only because the experimenter does not have the time and energy, but because many contingencies are too subtle and precise to be arranged without instrumental help. The same problem arises in education.

Consider, for example, the temporal patterning of behavior called "rhythm." Behavior is often effective only if properly timed. Individual differences in timing, ranging from the most awkward to the most skillful performances, affect choice of career and of artistic interests and participation in sports and crafts. Presumably a "sense of rhythm" is worth teaching, yet practically nothing is now done to arrange the necessary contingencies of reinforcement. The skilled typist, tennis player, lathe operator, or musician is, of course, under the influence of reinforcing mechanisms which generate subtle timing, but many people never reach the point at which these natural contingencies can take over.

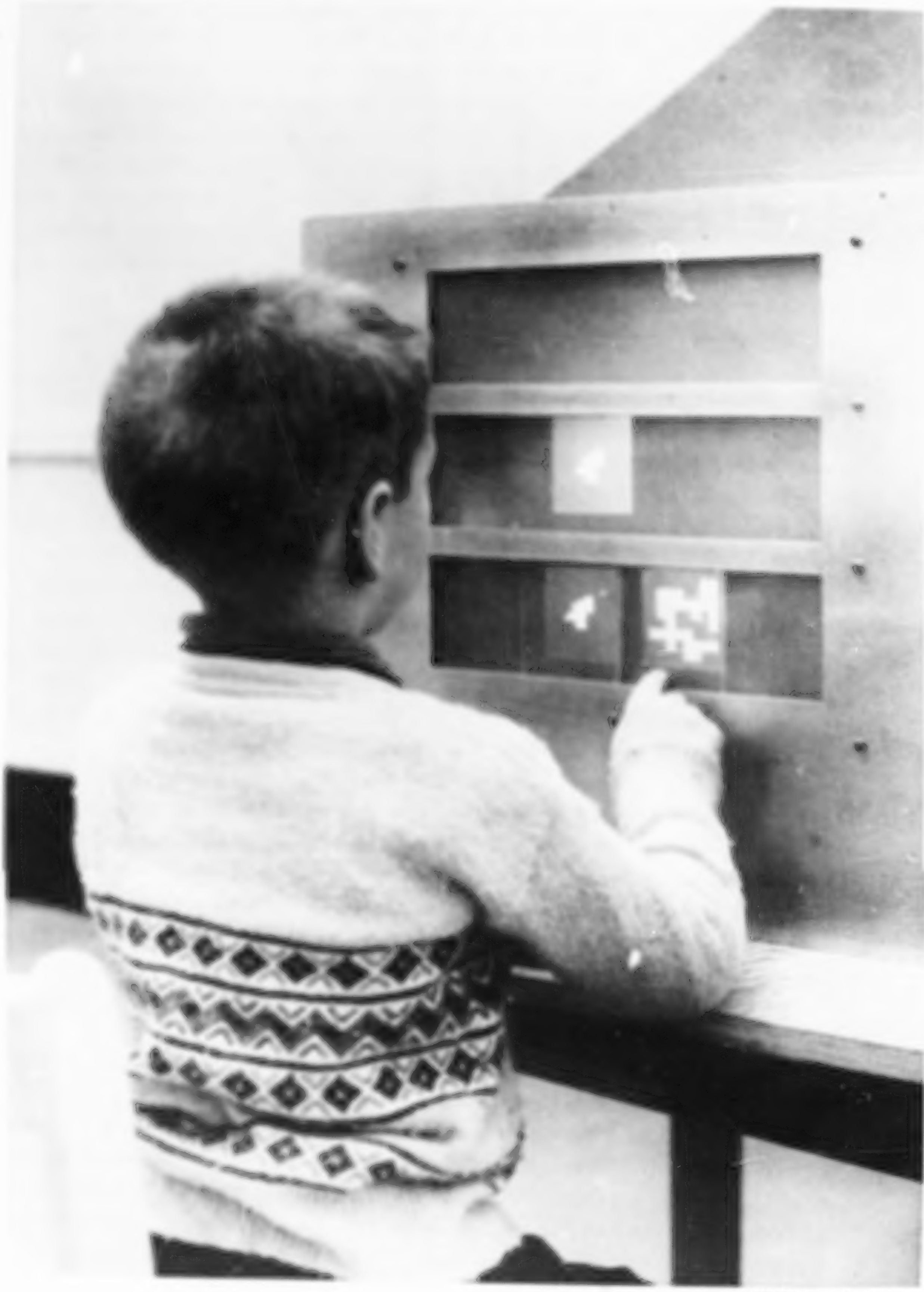
A relatively simple device supplies the necessary contingencies. The student taps a rhythmic pattern in unison with the device. "Unison" is specified very loosely at first (the student can be a little early or late at each tap) but the specifications are slowly sharpened. The process is repeated for various speeds and patterns. In another arrangement, the student echoes rhythmic patterns sounded by the machine, though not in unison, and again the specifications for an accurate reproduction are progressively sharpened. Rhythmic patterns can also be brought under the control of a printed score.

Another kind of teaching machine generates sensitivity to properties of the environment. We call an effective person "discriminating." He can tell the difference between the colors, shapes, and sizes of objects, he can identify three-dimensional forms seen from different aspects, he can find patterns concealed in other patterns, he can identify pitches, intervals, and musical themes and distinguish between different tempos and rhythms—and all of this in an almost infinite variety. Subtle discriminations of this sort are as important in science and industry and in everyday life as in identifying the school of a painter or the period of a composer.

The ability to make a given kind of discrimination can be taught. A pigeon, for example, can be *made* sensitive to the color, shape, and size of objects, to pitches, and rhythms, and so on—simply by reinforcing it when it responds in some arbitrary way to one set of stimuli and extinguishing responses to all others. The same kinds of contingencies of reinforcement are responsible for human discriminative behavior. *The remarkable fact is that they are quite rare in the environment of the average child.* True, children are encouraged to play with objects of different sizes, shapes, and colors, and are given a passing acquaintance with musical patterns; but they are seldom exposed to the precise contingencies needed to build subtle discriminations. It is not surprising that most of them move into adulthood with largely undeveloped "abilities."

The number of reinforcements required to build discriminative behavior in the population as a whole is far beyond the capacity of teachers. Too many teachers would be needed, and many contingencies are too subtle to be mediated by even the most skillful. *Yet relatively simple machines will suffice.* The apparatus shown in Figure 1 is adapted from research on lower organisms. It teaches an organism to discriminate selected properties of stimuli while "matching to sample." Pictures or words are projected on translucent windows which respond to a touch by closing circuits. A child can be made to "look at the sample" by reinforcing him for pressing the top window. An adequate reinforcement for this response is simply the appearance of material in the lower windows, from which a choice is to be made.

The child identifies the material which corresponds to the sample in some prescribed way by pressing one of the lower windows, and he is then reinforced again—possibly simply because a new set of materials now appears



Will Rapport

Fig. 1. A machine to teach the matching of colors, shapes, sizes, as well as correspondences between pictures and words, words and other words, and so on.

on the windows. If he presses the wrong window, all three choices disappear until the top window has been pressed again—which means until he has again looked at the sample. Many other arrangements of responses and reinforcements are, of course, possible. In an auditory version, the child listens to a sample pattern of tones and then explores other samples to find a match.

If devices similar to these were generally available in our nursery schools and kindergartens, our children would be far more skillful in dealing with their environments. They would be more productive in their work, more sensitive to art and music, better at sports, and so on. They would lead more effective lives. We cannot assert all this with complete confidence on the present evidence, but there is no doubt whatsoever *that the conditions needed to produce such a state of affairs are now lacking.* In the light of what we know about differential contingencies of reinforcement, the world of the young child is shamefully impoverished. And only machines will remedy this, for the required frequency and subtlety of reinforcement cannot otherwise be arranged.

The teacher is, of course, at a disadvantage in teaching skilled and discriminative behavior because such instruction is largely nonverbal. It may be that the methods of the classroom, in which the teacher is said to "communicate" with the student, to "impart information," and to build "verbal abilities," are better adapted to standard subjects matters, the learning of which is usually regarded as more than the acquisition of forms of behavior or of environmental control. Yet a second look may be worthwhile. Traditional characterizations of verbal behavior raise almost insuperable problems for the teacher, and a more rigorous analysis suggests another possibility. We can define terms like "information," "knowledge," and "verbal ability" by reference to the behavior from which we infer their presence. *We may then teach the behavior directly.* Instead of "transmitting information to the student" we may simply set up the behavior which is taken as a sign that he possesses information. Instead of teaching a "knowledge of French" we may teach the behavior from which we infer such knowledge. Instead of teaching "an ability to read" we may set up the behavioral repertoire which distinguishes the child who knows how to read from one who does not.

To take the last example, a child reads or "shows that he knows how to read" by exhibiting a behavioral repertoire of great complexity. He finds a letter or word in a list on demand; he reads aloud; he finds or identifies objects described in a text; he rephrases sentences; he obeys written instructions; he behaves appropriately to described situations; he reacts emotionally to described events; and so on, in a long list. He does none of this before learning to read and all of it afterwards. To bring about such a change is an extensive assignment, and it is tempting to try to circumvent it by teaching something called "an ability to read" from which all these specific behaviors will flow. But this has never actually been done. "Teaching reading" is always directed toward setting up specific items in such a repertoire.

It is true that parts of the repertoire are not independent. A student may acquire some kinds of responses more readily for having acquired others, and he may for a time use some in place of others (for example, he may follow written directions not by responding directly to a text but by following his own spoken instructions as he reads the text aloud). In the long run all parts of the repertoire tend to be filled in, not because the student is rounding out an ability to read, but simply because all parts are in their several ways useful. They all continue to be reinforced by the world at large after the explicit teaching of reading has ceased.

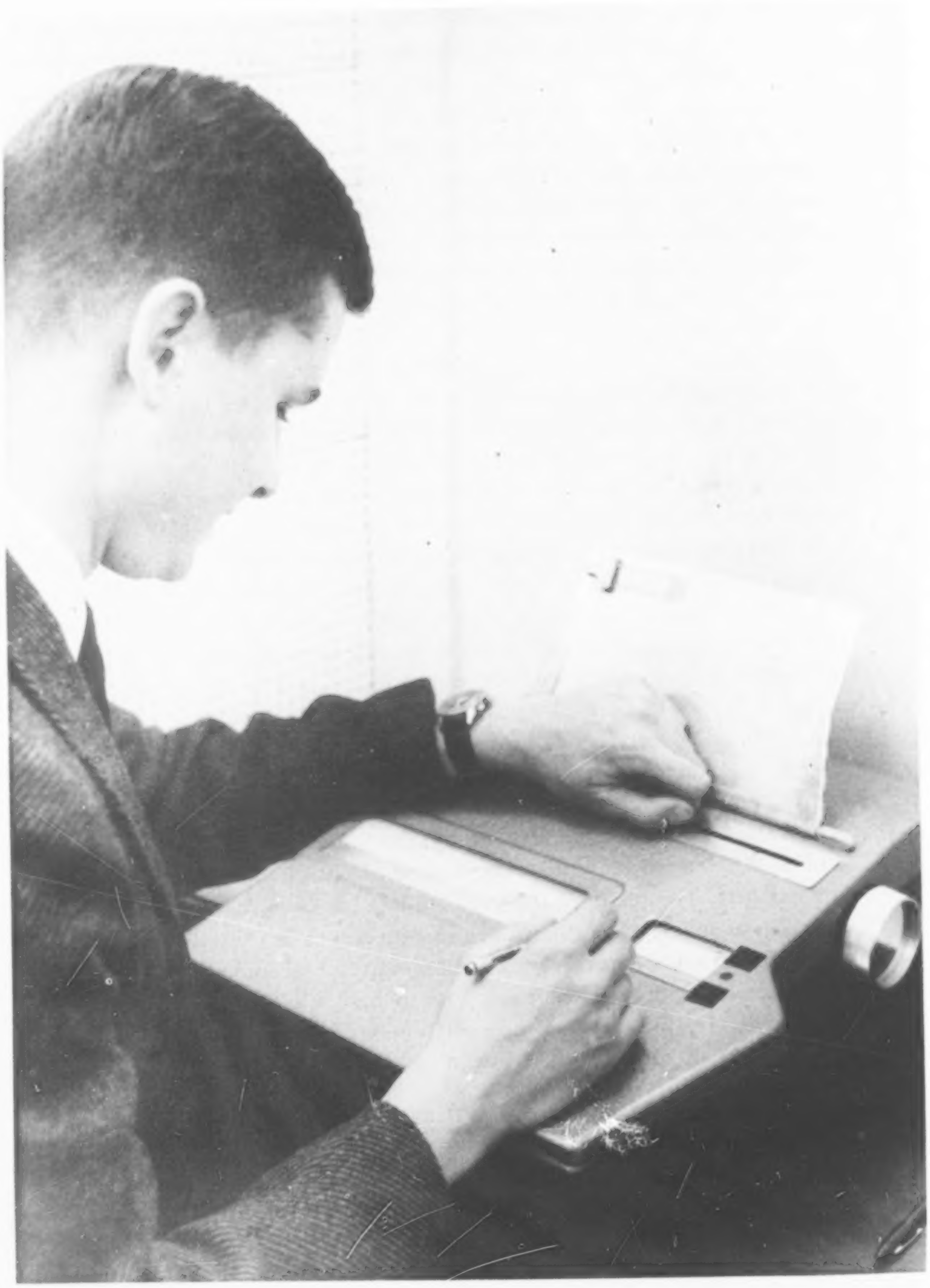
Viewed in this way, reading can also be most effectively taught with instrumental help. A pupil can learn to distinguish among letters and groups of letters in an alphabet simply as visual patterns in using the device and procedures just described. He can be taught to identify arbitrary correspondences (for example, between capitals and lower-case letters, or between handwritten and printed letters) in a more complex type of stimulus control which is within reach of the same device. With a phonographic attachment, correspondences between printed letters and sounds, between sounds and letters, between words and sounds, between sounds and printed words, and so on, can be set up. (The student could be taught all of this without pronouncing a word, and it is possible that he would learn good pronunciation more quickly if he had first done so.)

The same device can teach correspondences between words and the properties of objects. The pupil selects a printed or spoken word which corresponds in the language to, say, a pictured object or another printed or spoken word. These semantic correspondences differ in important respects from formal matches, but the same processes of programming and reinforcement can—indeed, must—be used. Traditional ways of teaching reading establish all these repertoires, but they do so indirectly and, alas, inefficiently. In "building a child's need to read," in motivating "his mental readiness," in "sharing information," and so on, the teacher arranges, sometimes almost surreptitiously, many of the contingencies just listed, and these are responsible for whatever is learned. An explicit treatment clarifies the program, suggests effective procedures, and guarantees a coverage which is often lacking with traditional methods. Much of what is called reading has not been covered, of course, but it may not need to be taught, for once these basic repertoires have been established, the child begins to receive automatic reinforcement in responding to textual material.

The same need for a behavioral definition arises in teaching other verbal skills (for example, a second language) as well as the traditional subjects of education. In advancing to that level, however, we must transcend a limitation of the device in Figure 1. The student can *select* a response without being able to speak or write, but we want him to learn to *emit* the response, since this is the kind of behavior which he will later find most useful. The emission of verbal behavior is taught by another kind of machine. A frame of textual material appearing in the square opening is incomplete: in place of certain letters or figures there are holes. Letters or figures can be made to appear in these holes by moving sliders (a keyboard would be an obvious improvement). When the material has been completed, the student checks his response by turning a crank. The machine senses the settings of the sliders and, if they are correct, moves a new frame of material into place, the sliders returning to their home position. If the response is wrong, the sliders return home, and a second setting must be made.

The machine can tell the student he is wrong without telling him what is right. This is an advantage, but it is relatively costly. Moreover, correct behavior is rather rigidly specified. Such a machine is probably suitable only for the lower grades. A simpler and cheaper procedure, with greater flexibility, is to allow the student to compare his written response with a revealed text. The device shown in Figure 2 uses this principle. It is suitable for verbal instruction beyond the lower primary grades—that is, through junior high school, high school, and college, and in industrial and professional education. Programmed material is stored on fan-folded paper tapes. One frame of material, the size of which may be varied with the nature of the material, is exposed at a time. The student writes on a separate paper strip. He cannot look at unauthorized parts of the material without recording the fact that he has done so, because when the machine has been loaded and closed, it can be opened only by punching the strip of paper.

The student sees printed material in the large window at the left. This may be a sentence to be completed, a question to be answered, or a problem to be solved. He writes his response in an uncovered portion of a paper strip at the right. He then moves a slider which covers the response he has written with a transparent mask and uncovers additional material in the larger opening. This may tell him that his response is wrong without telling him what is right. For example, it may list a few of the commonest errors. If the response he wrote is among them, he can try again on a newly uncovered portion of the paper strip. A further operation of the machine covers his second attempt and uncovers the correct response. The student records a wrong response by punching a hole alongside it, leaving a record for the instructor who may wish to review a student's performance, and operating a counter which becomes visible at the end of the set. Then the student records the number of mistakes he has made and may compare it with a par score for the set.



*Will Rapport*

FIG. 2. A machine to teach "verbal knowledge."

Exploratory research in schools and colleges indicates that what is now taught by teacher, textbook, lecture, or film can be taught in half the time with half the effort by a machine of this general type.[1] One has only to see students at work to understand why this is a conservative estimate. The student remains active. If he stops, the program stops (in marked contrast with classroom practice and educational television); but there is no compulsion for he is not inclined to stop. Immediate and frequent reinforcement sustains a lively interest. (The interest, incidentally, outlasts any effect of novelty. Novelty may be relevant to interest, but the material in the machine is always novel.) Where current instructional procedures are highly efficient, the gain may not be so great. In one experiment[2] involving industrial education there was approximately a 25% saving in the time required for instruction, something of the order of a 10% increase in retention, and about 90% of the students preferred to study by machine. In general, the student generally likes what he is doing; he makes no effort to escape—for example, by letting his attention wander. He need not force himself to work and is usually free of the feeling of effort generated by aversive control. He has no reason to be anxious about impending examinations, for none are required. Both he and his instructor know where he stands at all times.

No less important in explaining the success of teaching machines is the fact that each student is free to proceed at his own rate. Holding students together for instructional purposes in a class is probably the greatest source of inefficiency in education. Some efforts to mechanize instruction have missed this point. A language laboratory controlled from a central console presupposes a group of students advancing at about the same rate, even though some choice of material is permitted. Television in education has made the same mistake on a colossal scale. A class of twenty or thirty students moving at the same pace is inefficient enough, but what must we say of all the students in half a dozen states marching in a similar lock step?

In trying to teach more than one student at once we harm both fast and slow learners. The plight of the good student has been recognized, but the slow learner suffers more disastrous consequences. The effect of pressure to move beyond one's natural speed is cumulative. The student who has not fully mastered a first lesson is less able to master a second. His ultimate failure may greatly exaggerate his shortcoming; a small difference in speed has grown to an immense difference in comprehension. Some of those most

[1] Under the direction of Allen Calvin of Hollins College, an 8th grade class in the Roanoke School System completed all the work of a 9th grade class in algebra in one term. Test scores were comparable with a normal 9th grade performance, and a test nine months later showed a retention of at least 90% of the material learned.

[2] More recent results with the same material improved in the light of the earlier experiment were reported by J. L. Hughes and W. J. McNamara at the Annual Meeting of the American Psychological Association in New York, September, 1961. Their work concerned the use of programmed texts in industrial education.

active in improving education have been tempted to dismiss slow students impatiently as a waste of time, but it is quite possible that many of them are capable of substantial, even extraordinary, achievements if permitted to move at their own pace. Many distinguished scientists, for example, have appeared to think slowly.

One advantage of individual instruction is that the student is able to follow a program without breaks or omissions. A member of a class moving at approximately the same rate cannot always make up for absences, and limitations of contact time between student and teacher make it necessary to abbreviate material to the point at which substantial gaps are inevitable. Working on a machine, the student can always take up where he left off or, if he wishes, review earlier work after a longer absence. The coherence of the program helps to maximize the student's success, for by thoroughly mastering one step he is optimally prepared for the next. Many years ago, in their *Elementary Principles of Education*[3], Thorndike and Gates considered the possibility of a book "so arranged that only to him who had done what was directed on page one would page two become visible, and so on." With such a book, they felt, "much that now requires personal instruction could be managed by print." The teaching machine is, of course, such a book.

In summary, then, machine teaching is unusually efficient because (1) the student is frequently and immediately reinforced, (2) he is free to move at his natural rate, and (3) he follows a coherent sequence. These are the more obvious advantages, and they may well explain current successes. But there are more promising possibilities: the conditions arranged by a good teaching machine make it possible to apply to education what we have learned from laboratory research and to extend our knowledge through rigorous experiments in schools and colleges.

The conceptions of the learning process which underlie classroom practices have long been out of date. For example, teachers and textbooks are said to "impart information." They expose the student to verbal and nonverbal material and call attention to particular features of it, and in so doing they are said to "tell the student something." In spite of discouraging evidence to the contrary, it is still supposed that if you tell a student something, he then knows it. In this scheme, teaching is the transmission of information, a notion which, through a false analogy, has acquired undue prestige from communication engineering. Something is undoubtedly transmitted by teacher to student, for if communication is interrupted, instruction ceases; but the teacher is not merely a source from which knowledge flows into the student. We cannot necessarily improve instruction by altering the conditions of transmission—as, for example, by changing to a different sensory modality. This is a mistake made by some so-called teaching machines which, accepting our failure to teach reading, have tried to restore communication by using recorded speech. The student no longer pores over a book, as in the traditional portrait; he stares into space with earphones on his head. For the same reasons improvements in the coding of information may not be immediately relevant.

---

[3] Thorndike, Edward & Gates, Arthur. *Elementary Principles of Education*. (New York: B. Macmillan Co., 1929).

The student is more than a receiver of information. He must take some kind of action. The traditional view is that he must "associate." The stream of information flowing from teacher to student contains pairs of items which, being close together or otherwise related, become connected in the student's mind. This is the old doctrine of the association of ideas, now strengthened by a scientific, if uncritical, appeal to conditioned reflexes: two things occurring together in experience somehow become connected so that one of them later reminds the student of the other. The teacher has little control over the process except to make sure that things occur together often and that the student pays attention to them—for example, by making the experiences vivid or, as we say, memorable. Some devices called teaching machines are simply ways of presenting things together in ways which attract attention. The student listens to recorded speech, for example, while looking at pictures. The theory is that he will associate these auditory and visual presentations.

But the action demanded of the student is not some sort of mental association of contiguous experiences. It is more objective and, fortunately, more controllable than that. To acquire behavior, *the student must engage in behavior*. This has long been known. The principle is implied in any philosophy of "learning by doing." But it is not enough simply to acknowledge its validity. Teaching machines provide the conditions needed to apply the principle effectively.

Only in the early stages of education are we mainly interested in establishing *forms* of behavior. In the verbal field, for example, we teach a child to speak, eventually with acceptable accent and pronunciation, and later to write and spell. After that, topography of behavior is assumed; the student can speak and write and must now learn to do so appropriately—that is, he must speak or write in given ways under given circumstances. How he comes to do so is widely misunderstood. Education usually begins by establishing so-called formal repertoires. The young child is taught to "echo" verbal behavior in the sense of repeating verbal stimuli with reasonable accuracy. A little later he is taught to read—to emit verbal behavior under the control of textual stimuli. These and other formal repertoires are used in later stages of instruction to evoke new responses without "shaping" them.

In an important case of what we call instruction, control is simply transferred from so-called formal to thematic stimuli. When a student learns to memorize a poem, for example, it is clearly inadequate to say that by reading the poem he presents to himself its various parts contiguously and then associates them. He does not simply read the poem again and again until he knows it. (It is possible that he could never learn the poem in that way.) Something else must be done, as anyone knows who has memorized a poem from the text. The student must make tentative responses while looking away from the text. He must glance at the text from time to time to provide fragmentary help in emitting a partially learned response. If a recalled passage makes sense, it may provide its own automatic confirmation, but if the passage is fragmentary or obscure, the student must confirm the correctness of an emitted response by referring to the text after he has emitted it.

A teaching machine facilitates this process. It presents the poem line by line and asks the student to read it. The text is then "vanished"—that is, it becomes less and less clear or less and less complete in subsequent presentations. Other stimuli (arising from the student's own behavior in this case) take over. In one procedure a few unimportant letters are omitted in the first presentation. The student reads the line without their help and indicates his success by writing down the omitted letters, which are confirmed by the machine. More of the line is missing when it again appears, but because he has recently responded to a fuller text, the student can nevertheless read it correctly. Eventually, no textual stimulus remains, and he can "recite" the poem.

(If the reader wishes to try this method on a friend or member of his family without a machine, he may do so by writing the poem on a chalk board in a clear hand, omitting a few unimportant letters. He should ask his subject to read the poem aloud but to make no effort to memorize it. He should then erase another selection of letters. He will have to guess at how far he can go without interfering with his subject's success on the next reading, but under controlled conditions this could be determined for the average student quite accurately. Again the subject reads the poem aloud, making no effort to memorize, though he may have to make some effort to recall. Other letters are then erased and the process repeated. For a dozen lines of average material, four or five readings should suffice to eliminate the text altogether. The poem can still be "read.")

Memorized verbal behavior is a valuable form of knowledge which has played an important role in classical education. There are other, and generally more useful, forms in which the same processes are involved. Consider, for example, a labeled picture. To say that such an instructional device "tells the student the name of the pictured object" is highly elliptical—and dangerous if we are trying to understand the processes involved. Simply showing a student a labeled picture is no more effective than letting him read a poem. He must take some sort of action. As a formal stimulus, the label evokes a verbal response, not in this case in the presence of other verbal behavior on the part of the student, but in the presence of the picture. The control of the response is to pass from the label to the picture; the student is to give the name of the pictured object without reading it.

The steps taken in teaching with labeled pictures can also be arranged particularly well with a machine. Suppose we are teaching medical-school anatomy at the textbook level. Certain labeled charts represent what is to be learned in the sense that the student will eventually (1) give the names of indicated parts and describe relations among them and (2) be able to point to, draw, or construct models of parts, or relations among them, given their names. To teach the first of these, we induce the student to describe relations among the parts shown on a fully labeled chart. One effect of this is that he executes the verbal behavior at issue—he writes the names of the parts. More important, he does this while, or just after, looking at corresponding pictured details. He will be able to write the names again while looking at a chart which shows only incomplete names, possibly only initial letters. Finally, he will be able to supply the complete names of parts identified only by number on still another chart.. His verbal responses have passed from the control of textual stimuli to that of pictured anatomical details. Eventually, as he studies a cadaver, the control will pass to the actual anatomy of the human body. In this sense he then "knows the names of the parts of the body and can describe relations among them."

(The device shown in Figure 2 is designed to skip one or two steps in "vanishing" textual stimuli. A fully labeled chart may be followed by a merely numbered one. The student writes the name corresponding to a number in the first space. If he cannot do this, he operates the machine to uncover, not merely some indication that he is right or wrong, but additional help—say, a few letters of the correct response.)

Learning a poem or the names of pictured objects is a relatively straightforward task.. More complex forms of knowledge require other procedures. At an early point, the main problem becomes that of analyzing knowledge. Traditionally, for example, something called a "knowledge of French" is said to permit the student who possesses it to do many things. One who possesses it can (1) repeat a French phrase with a good accent, (2) read a French text in all the senses of reading listed above, (3) take dictation in French, (4) find a word spoken in French on a printed list, (5) obey instructions spoken in French, (6) comment in French upon objects or events, (7) give orders in French, and so on. If he also "knows English," he can give the English equivalents of French words or phrases or the French equivalents of English words or phrases.

The concept of "a knowledge of French" offers very little help to the would-be teacher. As in the case of reading, we must turn to the behavioral repertoires themselves, for these are all that have ever been taught when education has been effective. The definition of a subject matter in such terms may be extraordinarily difficult. Students who are "competent in first-year college physics," for example, obviously differ from those who are not—but in what way? Even a tentative answer to that question should clarify the problem of teaching physics. It may well do more. In the not-too-distant future much more general issues in epistomology may be approached from the same direction. It is possible that we shall fully understand the nature of knowledge only after having solved the practical problems of imparting it.

Until we can define subject matters more accurately and until we have improved our techniques of building verbal repertoires, writing programs for teaching machines will remain something of an art. This is not wholly satisfactory, but there is some consolation in the fact that an impeccable authority on the excellence of a program is available. The student himself can tell the programmer where he has failed. By analyzing the errors made by even a small number of students in a pilot study, it is usually possible to work a great improvement in an early version of a program. (The machine shown in Figure 2 is designed to supply the necessary feedback to the programmer in a convenient form. When a student punches an error, he marks the back of the printed material, which eventually carries an item-by-item record of the success or failure of the programmer. This is obviously valuable during the experimental stages of programming, but it will also be desirable when machines are widely used in schools and colleges, since publishers can then periodically call in programs to be studied and improved by their authors. The information supplied might be compared to a record showing the percentage of students who have misunderstood each sentence in a text.)

The teaching machine shown in Figure 2 falls far short of the "electronic classrooms" often visualized for the schools and colleges of the future. Many of these, often incorporating small computers, are based on misunderstandings of the learning process. They are designed to duplicate current classroom conditions. When instruction is badly programmed, a student often goes astray, and a teacher must come to his rescue. His mistakes must be analyzed and corrected. This may give the impression that instruction is largely a matter of correcting errors. If this were the case, an effective machine would, indeed, have to follow the student into many unprofitable paths and take remedial action. But under proper programming nothing of this sort is required. It is true that a relatively important function of the teacher will be to follow the progress of each student and to suggest collateral material which may be of interest, as well as to outline further studies, to recommend changes to programs of different levels of difficulty, and so on, and to this extent a student's course of study will show "branching." But changes in level of difficulty or in the character of the subject need not be frequent and can be made as the student moves from one set of material to another.

Teaching machines based on the principle of "multiple choice" also often show a misunderstanding of the learning process. When multiple-choice apparatuses were first used, the organism was left to proceed by "trial and error." The term does not refer to a behavioral process but simply to the fact that contingencies of reinforcement were left to chance: some responses happened to be successful and others not. Learning was not facilitated or accelerated by procedures which increased the probability of successful responses. The results, like those of much classroom instruction, suggested that errors were essential to the learning process. But when material is carefully programmed, both subhuman and human subjects can learn while making few errors or even none at all. Recent research by Herbert S. Terrace,[8] for example, has shown that a pigeon can learn to discriminate colors practically without making mistakes. The control exerted by color may be passed, *via* a vanishing technique, to more difficult properties of stimuli—again without error. Of course we learn something from our mistakes—for one thing, we learn not to make them again—but we *acquire* behavior in other ways.

The teaching machines of S. J. Pressey,[9] the first psychologist to see the "coming industrial revolution in education," were mechanical versions of self-scoring test forms, which Pressey and his students also pioneered. They were not designed for programmed instruction in the present sense. The student was presumed to have studied a subject before coming to the machine. By testing himself, he consolidated what he had already partially learned. For this purpose a device which evaluated the student's selection from an array of multiple-choice items was appropriate. For the same purpose multiple-choice material can, of course, be used in all the machines described above. But several advantages of programmed instruction are lost when such material is used in straightforward instruction.

In the first place, the student should *construct* rather than *select* a response, since this is the behavior he will later find useful. Secondly, he should advance to the level of being able to emit a response rather than merely recognize a given response as correct. This represents a much more considerable achievement, as the difference between the sizes of reading and writing vocabularies in a foreign language demonstrates. Thirdly, and more important, multiple-choice material violates a basic principle of good programming by inducing the student to engage in erroneous behavior. Those who have written multiple-choice tests know how much time, energy, and ingenuity are needed to construct plausible wrong answers. (They must be plausible or the test will be of little value.) In a multiple-choice *test*, they may do no harm, since a student who has already learned the right answer may reject wrong answers with ease and possibly with no undesirable side-effects. The student who is *learning*, however, can scarcely avoid trouble. Traces of erroneous responses survive in spite of the correction of errors or the confirmation of a right answer. In multiple-choice material designed to teach "literary appreciation," for example, the student is asked to consider three or four plausible paraphrases of a passage in a poem and to identify the most acceptable. But as the student reads and considers inacceptable paraphrases, the very processes which the poet himself used in making his poem effective are at work to destroy it. Neither the vigorous correction of wrong choices nor the confirmation of a right choice will free the student of the verbal and nonverbal associations thus generated.

---

[8] Terrace, Herbert S. Discrimination Learning With and Without Errors (unpublished Ph.D. Dissertation, Department of Psychology, Harvard University, 1961).

[9] Pressey, S. J. A simple apparatus which gives tests and scores—and teaches. *School and Society*, 1926, *23*, 373-376. (This article and other articles concerning teaching machines by S. J. Pressey are included in Lumsdaine, A. A. & Glaser, Robert (eds.) *Teaching Machines and Programmed Learning: A Source Book* (Washington, D.C.: National Education Association, 1960).

Scientific subjects offer more specific examples. Consider an item such as the following, which might be part of a course in high school physics:

*As the pressure of a gas increases, volume decreases. This is because:*

(a) *the space between the molecules grows smaller*
(b) *the molecules are flattened*
(c) etc. . . .

Unless the student is as industrious and as ingenious as the multiple-choice programmer, it will probably not have occurred to him that molecules may be flattened as a gas is compressed (within the limits under consideration). If he chooses item (b) and is corrected by the machine, we may say that he "has learned that it is wrong," but this does not mean that the sentence will never occur to him again. And if he is unlucky enough to select the right answer first, his reading of the plausible but erroneous answer will be corrected only "by implication"—an equally vague and presumably less effective process. In either case, he may later find himself recalling that "somewhere he has read that molecules are flattened when a gas is compressed." And, of course, somewhere he has.

Multiple-choice techniques are appropriate when the student is to learn to compare and choose. In forming a discrimination (as with the device shown in Figure 1), an organism must be exposed to at least two stimuli, one of which may be said to be wrong. Similarly, in learning to "troubleshoot" equipment there may be several almost equally plausible ways of correcting a malfunction. Games offer other examples. A given hand at bridge may justify several bids or plays, no one of which is wholly right and all the others wrong. In such cases, the student is to learn the most expedient course to be taken among a natural array of possibilities. This is not true in the simple acquisition of knowledge—particularly verbal knowledge—where the task is only rarely to discriminate among responses in an array. In solving an equation, reporting a fact of history, restating the meaning of a sentence, or engaging in almost any of the other behavior which is the main concern of education, the student is to *generate* responses. He may generate and reject, but only rarely will he generate a set of responses from which he must then make a choice.

It may be argued that machines which provide for branching and decision-making are designed to teach more than verbal repertoires—in particular, that they will teach thinking. There are strategies in choosing from an array, for example, which require kinds of behavior beyond the mere emission of correct responses. We may agree to this without questioning the value of knowledge in the sense of a verbal repertoire. (The distinction is not between rote and insightful learning, for programmed instruction is especially free of rote memorizing in the etymological sense of wearing down a path through repetition.) If an "idea" or "proposition" is defined as something which can be expressed in many ways, then it may be taught by teaching many of these "ways." What is learned is more likely to generalize to comparable situations than a single syntactical form, and generalization is what distinguishes so-called deeper understanding.

But not all thinking is verbal. There are, first of all, alternative, parallel nonverbal repertoires. The mathematician begins with a verbal problem and ends with a verbal solution, but much of his intervening behavior may be of a different nature. The student who learns to follow or construct a proof entirely by manipulating symbols may not engage in this kind of thinking. Similarly, a merely verbal knowledge of physics, as often seen in the student who has "memorized the text," is of little interest to the serious educator. Laboratories and demonstrations sometimes supply contingencies which build some nonverbal knowledge of physics. Special kinds of teaching machines could help, for machines are not only not confined to verbal instruction, they may well make it possible to reduce the emphasis on verbal communication between teacher and student.

A more clear-cut example of the distinction between verbal and nonverbal thinking is musical composition. The composer who "thinks musically" does more than perform on an instrument or enjoy music. He also does more than use musical notation. In some sense he "thinks" pitches, intervals, melodies, harmonic progressions, and so on. It should not surprise us that individuals differ greatly in their "abilities" to do this, since the necessary contingencies are in very short supply. One might attack the problem by setting up an explicit kinesthetic repertoire in which "thinking a pitch" takes the form of identifying a position on a keyboard. A device which arranges the necessary contingencies is under development. With its help we may discover the extent to which students can in general learn (and at what ages they can learn most effectively) to strike a key which produces a tone which has just been heard.

2

Similar devices might generate important forms of nonverbal mathematical behavior or the behavior exhibited, say, by an inventor conceiving of a device in three dimensions, as well as creative repertoires in other forms of art. Here is an extraordinary challenge to the technology of instrumentation.

There is another sense in which the student must learn to think. Verbal and nonverbal repertoires may prepare him to behave in effective ways, but he will inevitably face novel situations in which he cannot at first respond appropriately. He may solve such problems, not by exercising some mental ability, but by altering either the external situation or the relative probabilities of parts of his own repertoire. In this way he may increase the probability of an adequate response.

In this sense, thinking consists of a special repertoire which we may call self-management. For example, the student may alter the extent to which the environment affects him by "attending" to it in different ways. As one step in teaching thinking we must teach effective attending. The phrase "Pay attention!" is as common on the lips of teachers as "Open, please" on those of dentists—and for much the same reason: both phrases set up working conditions. The student may pay attention to avoid punishment and in doing so may learn to pay attention, but where aversive sanctions have been given up, teachers have resorted to attracting and holding attention. The techniques of the publication and entertainment industries are extensively invoked. Primers are usually decorated with colored pictures, and high school textbooks are sometimes designed to resemble picture magazines. Films dramatize subject matters in competition with noneducational films and television.

Attention which is captured by attractive stimuli must be distinguished from attention which is "paid." Only the latter must be learned. Looking and listening are forms of behavior, and they are strengthened by reinforcement. A pigeon can learn to match colors, for example, only if it "pays attention to them." The experimenter makes sure that it does so, not by attracting its attention, but by reinforcing it for looking. Similarly, a well-taught student pays attention to sentences, diagrams, samples of recorded speech and music, and so on, not because they are attractive but because something interesting occasionally happens *after* he has paid attention.

Most audio-visual devices fail to teach attention because they stimulate the student *before* he looks or listens closely. No matter how well a four-colored text or a dramatically filmed experiment in physics attracts attention, it prepares the student only for comics, advertising, picture magazines, television programs, and other material which is *interesting on its face.* What is wanted is an adult who, upon seeing a page of black-and-white text, will read it because it may *prove* interesting. Unfortunately, the techniques associated with captured and paid attention are incompatible. Whenever a teacher attracts the attention of a student, he deprives him of an opportunity to learn to pay attention. Teaching machines, with their control over the consequences of action, can make sure that paying attention will be effectively reinforced.

Another activity associated with thinking is studying—not merely looking at a text and reading it but looking and reading *for the sake of future action.* Suppose we show a child a picture and later, in the absence of the picture, reinforce him generously for correct answers to questions about it. If he has done nothing like this before, he will probably not be very successful. If we then show him another picture, he may begin to behave in a different way: he may engage in behavior which will increase the probability that he will later answer questions correctly. It will be to his advantage (and to ours as educators) if this kind of behavior is taught rather than left to chance. We teach a student "how to study" when we teach him to take notes, to rehearse his own behavior, to test himself, to organize, outline, and analyze, to look for or construct mnemonic patterns, and so on. Some of these behaviors are obvious, but others are of more subtle dimensions and admittedly hard to teach. Machines have an advantage in maintaining the contingencies required for indirect or mediated reinforcement.

Other aspects of thinking, including the solution of personal problems, can also be analyzed and directly programmed. This is not current practice, however. Students are most often "taught to think" simply by thrusting them into situations in which already established repertoires are inadequate. Some of them modify their behavior or the situation effectively and come up with solutions. They may have learned, but they have not necessarily been taught, how to think.

Logicians, mathematicians, and scientists have often tried to record and understand their own thinking processes, but we are still far from a satisfactory formulation of all relevant behaviors. Much remains to be learned about how a skillful thinker examines a situation, alters it, samples his own responses with respect to it, carries out specific verbal manipulations appropriate to it, and so on. It is quite possible that we cannot teach thinking adequately until all this has been analyzed. Once we have specified the behavior, however, we have no reason to suppose that it will then be any less adaptable to programmed instruction than simple verbal repertoires.

Teaching machines and the associated practices of programmed instruction will have proved too successful if their practical consequences are allowed to overshadow their promise for the future. We need teaching machines to help solve a very pressing problem, but we also need them to utilize our basic knowledge of human behavior in the design of entirely new educational practices.

Teaching machines are an example of the technological application of basic science. It is true that current machines might have been designed in the light of classroom experience and common sense, and that explanations of why they are effective can be paraphrased in traditional terms. The fact remains that more than half a century of the self-conscious examination of instructional processes had worked only moderate changes in educational practices. The laboratory study of learning provided the confidence, if not all the knowledge, needed for a successful instrumental attack on the *status quo*. Traditional views may not have been actually wrong, but they were vague and were not entertained with sufficient commitment to work substantial technological changes.

As a technology, however, education is still immature, as we may see from the fact that it defines its goals in terms of traditional achievements. Teachers are usually concerned with reproducing the characteristics and achievements of already educated men. When the nature of the human organism is better understood, we may begin to consider not only what man has already shown himself to be, but what he may become under carefully designed conditions. The goal of education should be nothing short of the fullest possible development of the human organism. An experimental analysis of behavior, carried out under the advantageous conditions of the laboratory, will contribute to progress toward that goal. So will practical experiments conducted in schools and colleges with the help of adequate instrumentation.

PRECEDING PAGE BLANK-NOT FILMED

# NEW DIRECTIONS IN TEACHING-MACHINE RESEARCH

James G. Holland

Since the publication of Skinner's article "The Science of Learning and the Art of Teaching" in 1954, programmed instruction has advanced at an accelerated pace which promises great improvement in education. These programs have usually followed the format dictated by a single type of machine. Such a machine presents printed material and provides for a student's response to small units of material. Nearly every academic subject matter has some program development underway in this manner. Such work commonly uses a standard, favorite text as the accepted definition of what should be taught. This type of expedient program is indeed an important contribution to immediate educational needs; but, although the author intends no depreciation of this important effort, it misses a possibly greater potential contribution of the new behavioral technology which is now ready to begin the analysis of fundamental skills and the development of techniques for establishing them.

It is true that workable principles have been developed for expedient programming and that these have, in turn, bred their own particular type of research effort, namely, hypothesis testing. With research talent already drained by expedient programming, it is a shame to see it further drained by the type of research being carried out. Such pedestrian problems as response mode and comparisons of "programmed

[Reprinted from J.E. Coulson (Ed.) PROGRAMMED LEARNING AND COMPUTER-BASED INSTRUCTION, New York: John Wiley, 1961.]

instruction" and "conventional methods" are investigated. Aside from the overgeneralization inevitable from such ill-defined variables, even these relatively uninteresting problems are inadequately treated. Despite the overwhelming recognition of the importance of quality in the program, we are seldom given information for evaluating the program. Indeed, the importance of this point has been illustrated by a demonstration (Holland, 1960) that the difference between using a program and simply reading the same completed statements can be abolished by such a minor change in the program as leaving less appropriate parts of the statement blank for the student to complete. Moreover, hypothesis-testing studies in teaching-machine research often use only 100 or 200 items which, for the variables studied, should and do yield negative results. In the face of these obvious difficulties, it is all the more surprising when the researchers eagerly accept the null hypothesis.

As important as expedient programming is, a real revolution in education can be brought about by an experimental analysis of techniques to establish fundamental skills. Techniques must be developed to teach people to speak and write correctly rather than simply to teach *about* English. Techniques must be developed to teach persons to behave in a foreign language without necessarily teaching them about the language and without teaching them English equivalences which interfere with eventual "thinking" in the language. Techniques must be developed to teach logical speech and action rather than merely to teach about a formal discipline that logicians have developed. Techniques must be developed to establish such fundamental skills as attentive reading, observation of displays for later recall, inductive reasoning, and a host of discriminative capacities. This effort will require the experience accumulated over the years in the behavioral laboratories. It is known that gradual progression, control of the student's observing or mediating behavior, and variation of material are all important general principles that should apply. But the specific form that the application will take will be an experimental matter to be worked out by careful experimental analysis. Such research attempts to discover conditions under which a skill can be created; it attempts to discover new methods and techniques. Work of this type is carried out with extreme care and with careful observation and sensitivity to all possible influences. The experimenter cannot be bound by an existing format, type of material, or existing machine in his endeavors to meet these goals.

## Discrimination Learning

Some of the possibilities of discrimination learning will serve to illustrate the potential of work on basic skills. Many of us cannot even see the things an artist sees in his paintings. Some of us cannot discriminate the tones and intervals so important to the musician. The connoisseur of wine is more than someone who knows the good vintage years as might be taught with a standard verbal program. The true connoisseur has acquired subtle taste discriminations. But, in addition to cases of simple differential responses to discriminative stimuli, discrimination training is important for the establishment of minimal repertories in which cases slight differences in stimuli serve automatically to reinforce behavior which generates them (cf. Holland & Skinner, 1961, pp. 157–174). Such repertories consist of a continuous field of responses which generates stimulus patterns also capable of continuous variation. Skilled handwriting or drawing from copy produces lines on paper which differ in degree of adequacy. It is only after an individual has learned to discriminate good handwriting that he can be automatically reinforced by the production of good letter patterns. Evans (1961) has shown that teaching young children to discriminate letter patterns vastly improves their ability to draw them without additional formal training on actual production of pattern. The discrimination training makes possible immediate, automatic, differential reinforcement for production of correct patterns. Such differential reinforcement is possible only after the discrimination has been acquired.

Young children (or for that matter adults) who have articulation disorders usually cannot discriminate the difference between the correct sound and the sound they produce for it. It is common for a speech therapist to begin teaching discrimination of the sound before working on actual articulation. Only after learning to hear the sound can the child receive the automatic differential reinforcement for correct production. This problem also applies to learning new languages; for example, the Korean learning English cannot even *hear* the difference between the words hit, hat, hot, and heat without *learning* to do so. The difficulty of many American students in pronouncing the French "eu" stems in part from a similar inability to discriminate that sound from the English "oo."

A technique has been developed by Audrey Holland and Jack Matthews (in preparation) to establish sound discriminations in 7-

and 8-year-old children requiring articulation therapy. In this case the teaching machine is a modified tape-recorder which presents spoken words. The child presses one of a set of buttons on hearing one of a pair of sounds. When he is correct, the tape recorder simply continues to play; when he is incorrect it rewinds a short distance and replays that item. The program illustrates gradual progression and proper variation of extraneous stimulus elements to get the (*s*) sound established as the discriminative stimulus; it also illustrates control in forcing very careful notice of the sounds. The program begins with the student pressing one button if an isolated sound is an (*s*) sound and another if it is any other sound. This phase begins with quite different sounds *s—ch—ch—s*; and gradually increases in difficulty until the sounds are very much more alike *s—sh—th—s—th*. In the next task the child listens to pairs of words and presses a button numbered "1" if the first word has an (*s*) sound and button "2" if the second word has an (*s*). Early items have an (*s*) sound in the beginning position (e.g., early in the series: *beautiful, Saturday;* later in the series: *Sue, zoo*); later items have (*s*) sounds in the final position (e.g., *ice, eyes; bush, hoops; race, raise*); and still later items have (*s*) sounds in the middle of the words—this being the most difficult position in which to discriminate the difference (e.g., from *worry, baseball* to *racer, razor*). This sequence of position moves from easy to more difficult, but within each of these the *discriminations* move also from easy to difficult because of the choices of words and because the recorded voice accentuates the differences where needed.

Next, the child is presented single words and presses one of three buttons to indicate whether the (*s*) sound occurred in the initial, medial, or final position of words. This is a highly important part of the program because it forces very careful observation of the (*s*), and it is the most difficult portion of the program.

In the final part of this program the child discriminates which of a pair of productions of the same word contains the correct (*s*) sound. Among the misarticulated alternatives there are equal representations for 14 possible misarticulations of (*s*).

The entire program includes 585 items and, in addition to following a gradual progression, illustrates very well the other principles. The correct stimulus control is assured by including many words which contain the (*s*) *sound* but which are not *spelled* with an (*s*) and words which are *spelled* with (*s*) but lack the (*s*) sound. Moreover, within the program the (*s*) sound precedes and follows all the vowel sounds;

the (*s*) sound occurs in all positions within the words used and, furthermore, the tasks vary—there is identification of isolated sounds; discrimination of words containing (*s*) sounds, discrimination of position of (*s*) sounds within words, and discrimination of correct from misarticulated (*s*) sounds within words.

When children with defective (*s*) discrimination and articulation worked through this program, their ability to discriminate (*s*) sounds improved drastically. In addition, even though there had been no work on their actual articulation, a week after completing the program there was a significant improvement in their (*s*) articulation.

## Inductive Reasoning

Some work currently underway by Eugene Long and the present writer illustrates an effort to develop procedures which may enable training of "mental abilities." The instrument used is shown in Fig. 1.

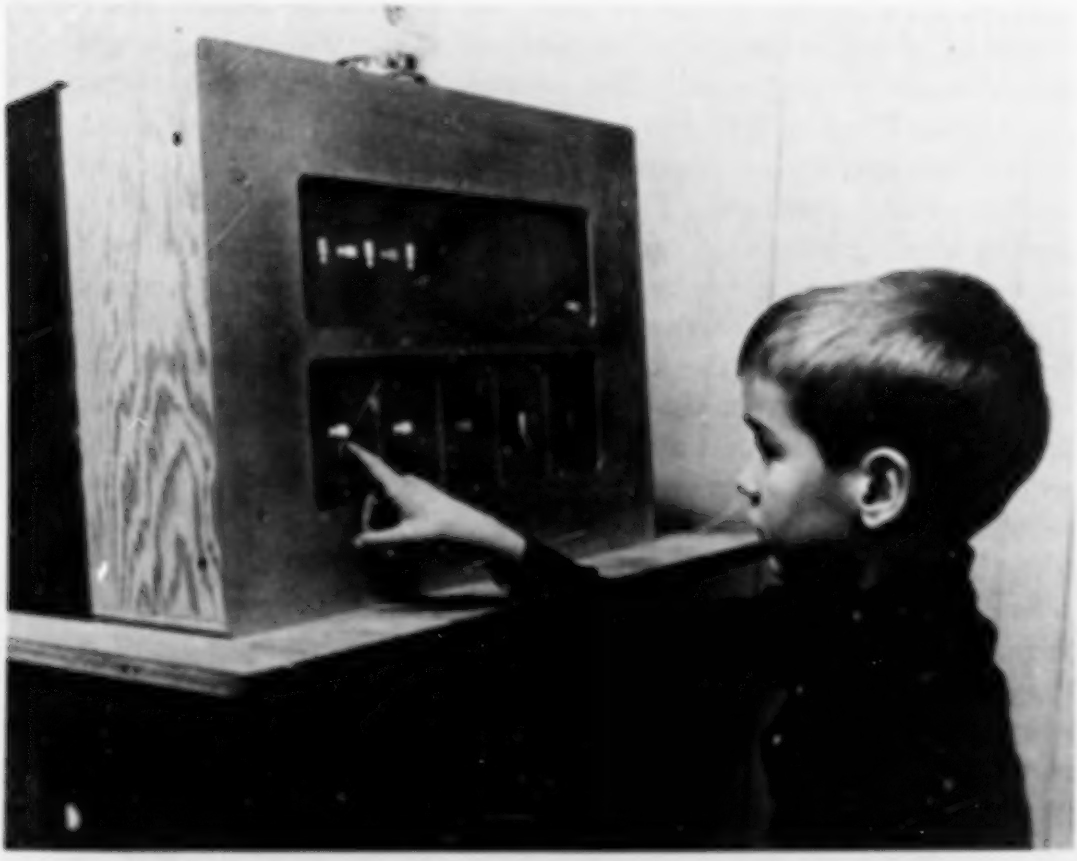


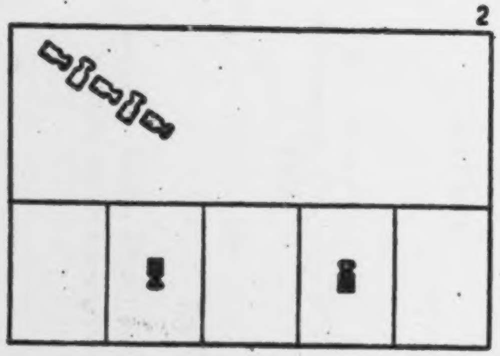
*Fig. 1. The subject's panel. The upper window displays the problem and the lower five windows the response alternatives. The globe at the top lights when a response is correct.*

There is a top window and five smaller choice windows which serve both as stimulus displays and response keys. A slide projector trans-illuminates these keys. When the top window is pressed, a shutter opens to expose the projected images on five small choice windows. Photocells in the projector, operated by a hole punch coding system, determine which of the five bottom windows will be scored as correct. When the student presses the bottom window that is correct for the problem presented in the top window, the shutters close, the slide changes presenting the next frame, and simultaneously a red light at the top of the instrument lights and a gong sounds. If, instead, the subject presses an incorrect window, only the bottom shutter closes. He must then press the top window, again opening the bottom shutter, and respond to the same problem once more. If he is now correct, the slide changes, the light goes on, and the gong sounds; but instead of presenting the next frame in the sequence, the projector presents the previous frame. If he is now correct on this frame, the material advances to the frame that he previously missed, and if he is then right on his very first choice the material advances to the next item in the sequence.

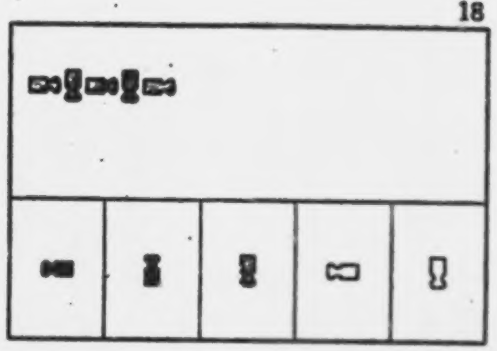
The details of this procedure were *not* arbitrarily decided upon. Considerable exploratory work carried out at the Walter E. Fernald State School for the Mentally Retarded enabled the evolution of this procedure. The red light and gong were added to provide clear-cut stimuli following correct responses. Candies (M & M's) were tried as reinforcers and discarded. Requiring a first correct response before moving ahead prevents the subject from getting through long, difficult sequences of the program by essentially chance or random choices. Thus the procedure decreases the danger of fortuitously reinforcing sequential patterns on the keys and other nonstimulus related responses. Moreover, when this procedure was not used, the student who had advanced by such fortuitous responding had difficulty later in the sequence even if he attempted to use the stimuli. This was because he was then faced with more advanced problems than he had seen when he initially began his random responding.

Samples of the material used to establish inductive reasoning are shown in Fig. 2. The child is presented with a sequence of forms in the problem window. The child's task is to press the bottom window which displays a form that could be added to extend the series correctly one element to the right. This material, incidentally, is somewhat similar to test material on inductive reasoning in which the

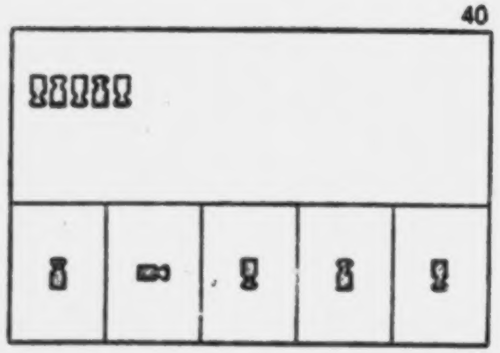
2



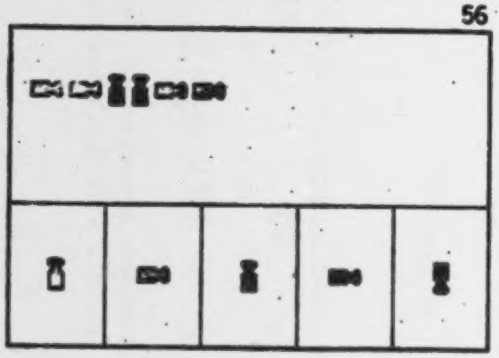
18



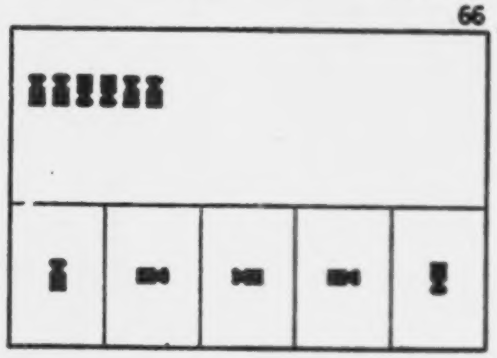
40



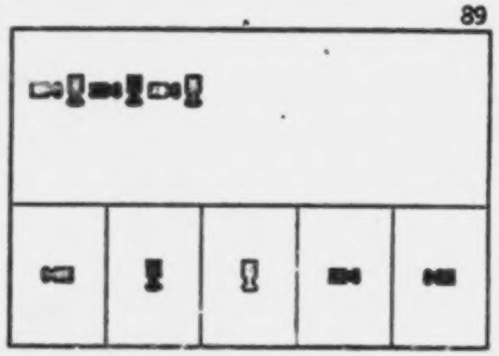
56



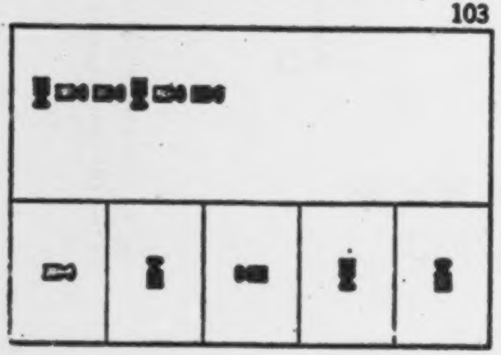
66



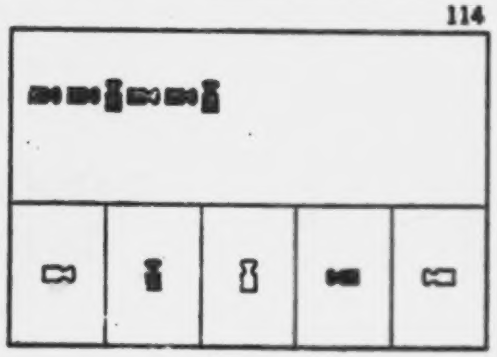
89



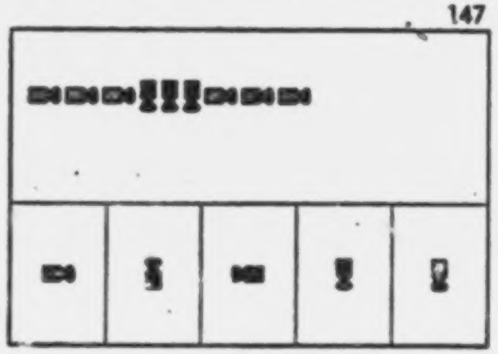
103



114



147



child's task is to determine what would follow next in a sequence. There are also a small number of very simple problems of this type and an ingenious set of booklets developed by Thelma Thurstone (1948) to teach primary mental abilities. The other frames shown in Fig. 2 are taken from various points throughout the program and the number above them indicates their position in the program. At present 234 items in this program carry the student to the level indicated by number 233 in Fig. 2. The last 31 frames of the program are of the general type illustrated by number 233 in Fig. 2, in which the correct alternative (window number 1) shows a sample not actually shown in the problem window because the problem sequence was not extended far enough to include it. The progression from simple to difficult problems is obvious in the figure. Within each type of problem a progression of difficulty is provided by making the incorrect alternatives progressively more difficult (contrast number 18 with number 40), and by occasionally decreasing from two to one the number of relative dimensions (direction and color) enabling solution.

The program as it now stands is in transition. Additional revisions will be made, with considerable extension into more difficult problems and different types of material. The data shown in Fig. 3 were obtained with an earlier version of the program consisting of 187

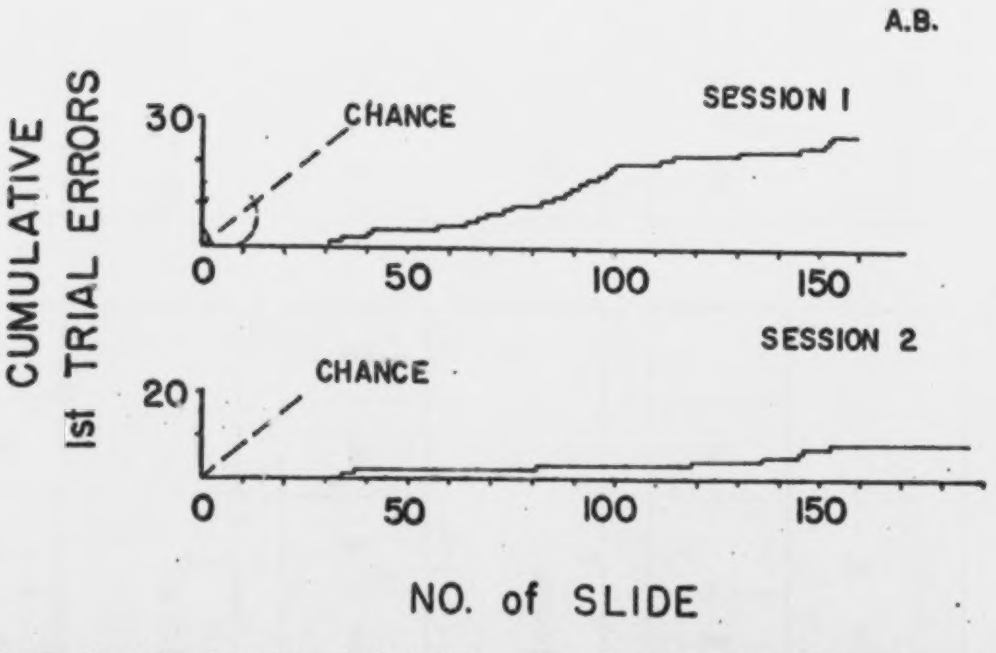


*Fig. 3. Cumulative first trial errors for a single subject. Session 1 is the subject's performance on the 187-item sequence. Session 2 is the same subject on a revised 234-item sequence.*

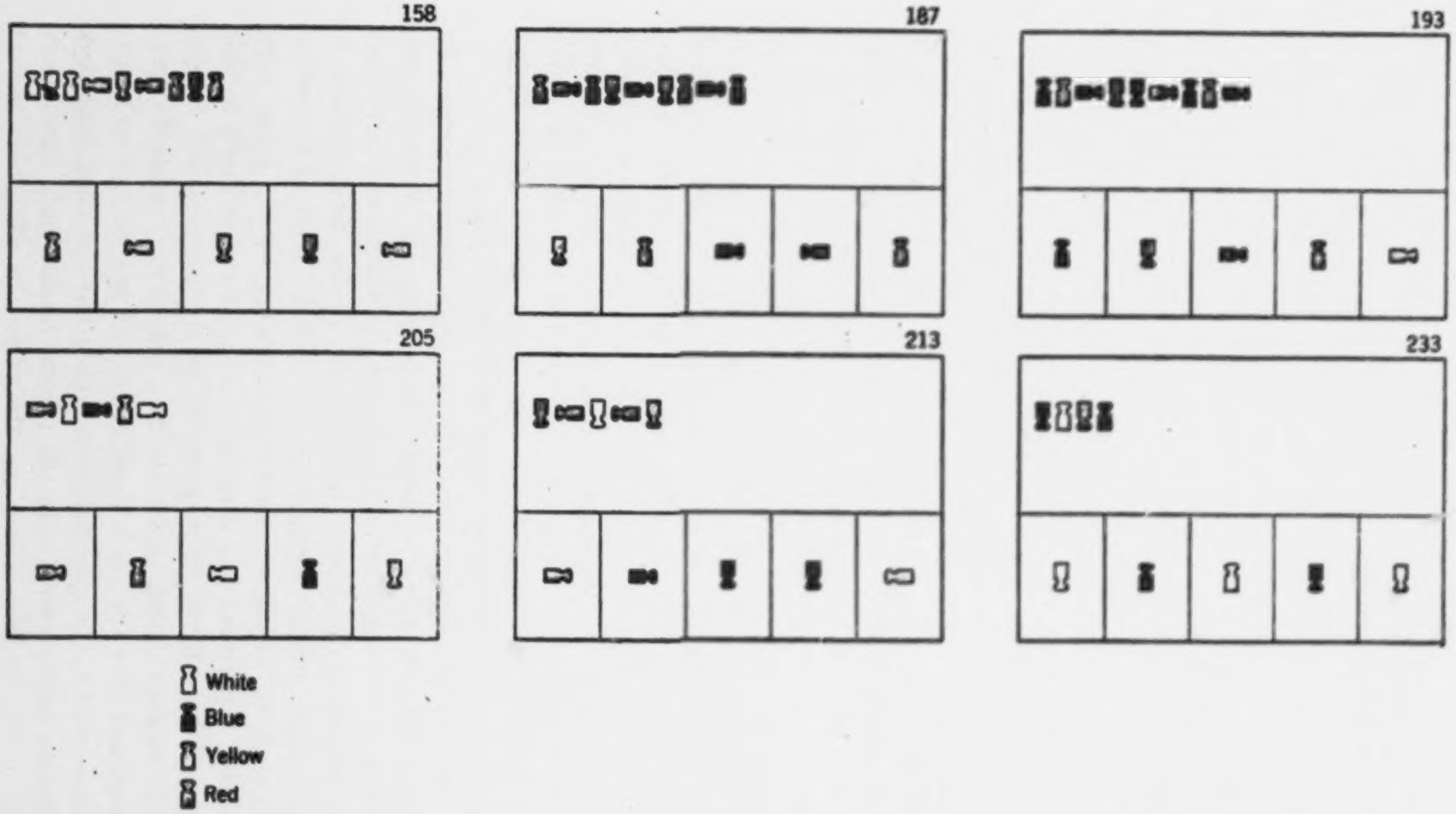

158
187
193
205
213
233
White
Blue
Yellow
Red

*Fig. 2. Sample items from the inductive reasoning experiment. The number above each item indicates its position in the 234-item sequence. The correct alternatives (in order) of these items are 4, 3, 3, 3, 5, 4, 4, 1, 4, 2, 1, 2, 4, 5, and 1.*

frames. The practice has been to try initial material and to modify it immediately from experience with a few subjects, try the revision and again modify it, and so forth until material has been developed to establish the behavior desired. In Fig. 3 there are two sessions shown for a single subject. The first session uses the 187-item version. The data shown are a plot of cumulative first-trial errors. On the horizontal axis the unit is the individual slide in its proper place in the sequence. When on an individual trial the slide is presented and an incorrect response is made, there is a step up of one unit. When another error is made on the presentation of another new slide, there is another upward step, resulting in a cumulative plot of first trial errors during the session. This means of presenting the data shows each area in which errors are clustered. The slope in any portion of the graph indicates error rate in that region. The line labeled "chance" is the slope that the record would take if the child's responses were under no stimulus control at all. In this subject's first session he made no errors until slide #31. An appreciable cluster of errors occurred in the region of slide #70 when the problem was double alternation in one dimension and single alternation in the other. It should be noted, nevertheless, that even in this area his performance was considerably better than chance. Again, another area of difficulty appeared around slide #90 when the problem was of the one-two-one nature; the errors between #90 and #100 resulted from inadequate bridging material from double to triple alternation, and the cluster of three errors near the end of the sequence occurred for the items in which the correct response was not directly seen in the sample (that is, the difficult problems like that shown as number 233 in Fig. 2).

On the basis of this performance, the sequence was improved. In no case were difficult slides removed. Improving the material consisted entirely of rearrangements in order and additions of other slides to provide smooth progression. Incidentally, in the final portion of the program, more difficult items were also added. Even if the material had not been changed, some improvement would have been expected. But experience has demonstrated that errors made in this kind of material are likely to be repeated in another session. Simple drill procedures with similar materials show only a very gradual reduction in errors and slow approach to criterion. At any rate, if no reduction had been shown, the experimenters would have learned that they had effected no improvement. In the second session only nine errors were made in the longer version and the places in which clusters of errors occurred previously had few errors. In what would have been the first cluster of six errors above only one error was made; in the third cluster, one error; and in the final cluster, one error.

The sequence of 234 frames developed was used with naïve, normal subjects ranging from 6 to 9 years of age. The results for a single session are shown in Fig. 4. One subject, D. W., age 9, made only two errors in the entire program, and those were in the most difficult material at the end. Subject B. M., age 6, made only nine errors in the entire sequence and only five of these were on the 31 difficult problems near the end of the program. J. D., age 7, made 11 errors. Interestingly only one of these was on the last 20 frames. This is possibly related to the two short rest periods allowed him because of signs of restlessness. Subject R. M., age 8, made 13 errors and remained considerably better than chance even near the end of the program where more than half of his errors occurred. D. W. and C. D., both 7, also did well during most of the program but revealed deficiencies near the end where they performed at near chance for the

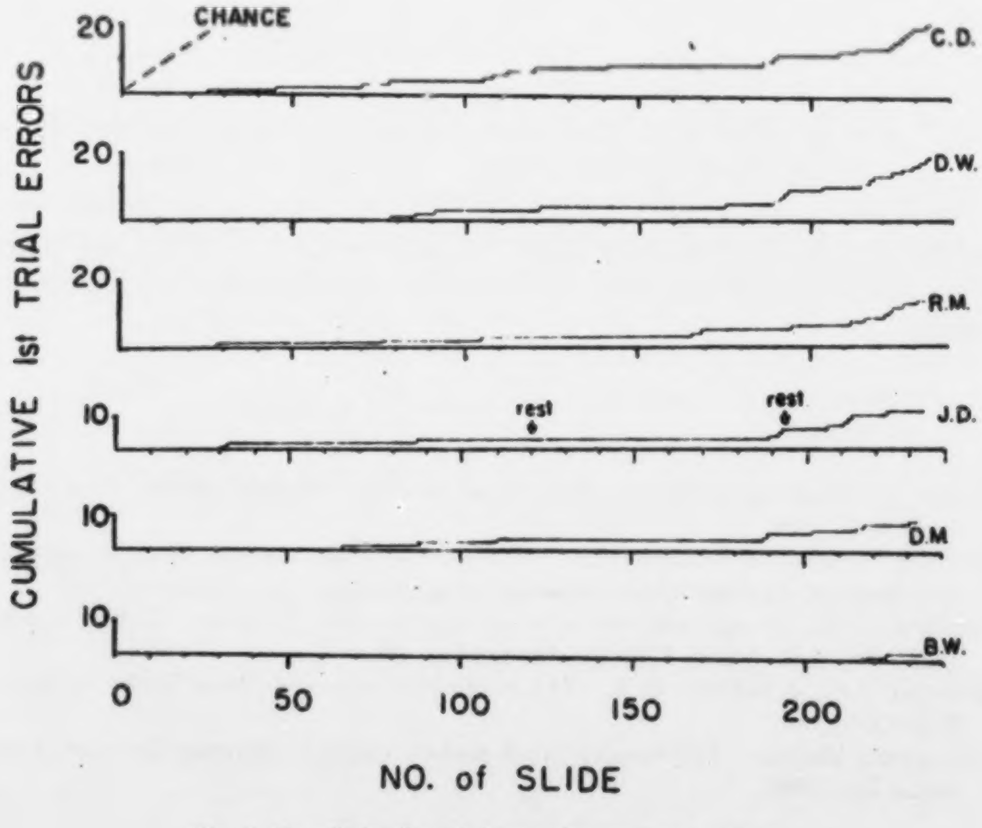


*Fig. 4. Cumulative first-trial errors for six naïve subjects.*

last 10 frames. Several of the subjects also revealed a deficiency in the material in the region of frames 185 to 195.

This research has successfully developed a procedure that will establish an ability to solve an inductive sequence of medium difficulty. The work, however, is not complete. Aside from the small revisions which must be made in view of the foregoing data, other types of problems need to be included and considerably more difficult material must be added. Beyond this phase it would be of interest to posttest with standardized abilities tests and to investigate further the limits of transfer.

This approach to learning research is novel in that it is directed toward development of techniques that will establish a given behavior. It can be contrasted with research directed toward discovering some supposed "learning function" or to discover whether one arbitrary condition is better than another arbitrary condition. Through careful analytical research of procedures, one can often attain behavioral control. The real product of such work is a technique and procedure for establishing such control rather than the usual confirmation of an hypothesis. One understands language behavior when he knows in detail how to produce it. One understands inductive reasoning when he knows how to create the ability. Although the interests of the investigators are primarily those of pure science, it is, of course, conceivable that the day may come when preschool children will play with learning material that will establish a wide range of important skills.

### REFERENCES

Evans, J. Teaching machines. Paper read at Amer. Psychol. Assoc., New York, September, 1961.

Holland, Audrey, & Matthews, J. The use of a teaching machine in training speech sound discrimination for articulation therapy. In preparation.

Holland, J. G. Design and use of a teaching machine program. Paper read at Amer. Psychol. Assoc., Chicago, September, 1960.

Holland, J. G., & Skinner, B. F. *The analysis of behavior.* New York: McGraw-Hill, 1961.

Thurstone, Thelma. *Learning to think series.* Chicago: Science Research Associates, Inc., 1948.

# A MULTIPLE - CHOICE VISUAL DISCRIMINATION APPARATUS[1]

**Wells Hively**

This apparatus was designed to program stimuli to teach complex visual discrimination tasks (Hively, 1962; Hively, 1964). It has been extensively used with pre-school and kindergarten children, and could be adapted for use with other primates.

An over-all view of the apparatus is shown in Fig. 1. Programming and recording circuitry are housed separately. An automatic slide projector projects stimuli onto the windows of the operandum. The subject responds to the stimuli by touching the windows. A switch beneath each window is wired in series with a photocell which may be activated by a spot of light from the projector. These are the basic elements of the apparatus.

Details of the operandum are shown in Figs. 2 and 3. The windows are quarter-inch plexiglass, coated with a commercial preparation for rear projection screens (Polacoat Inc., Blue Ash, Ohio). They are hung on steel pins as shown in Fig. 2, the diameter of the holes in the windows being just slightly larger than the diameter of the pins, allowing the windows to swing slightly. The bottom of each window rests against the pin of a microswitch. Pressing the window operates the switch, and

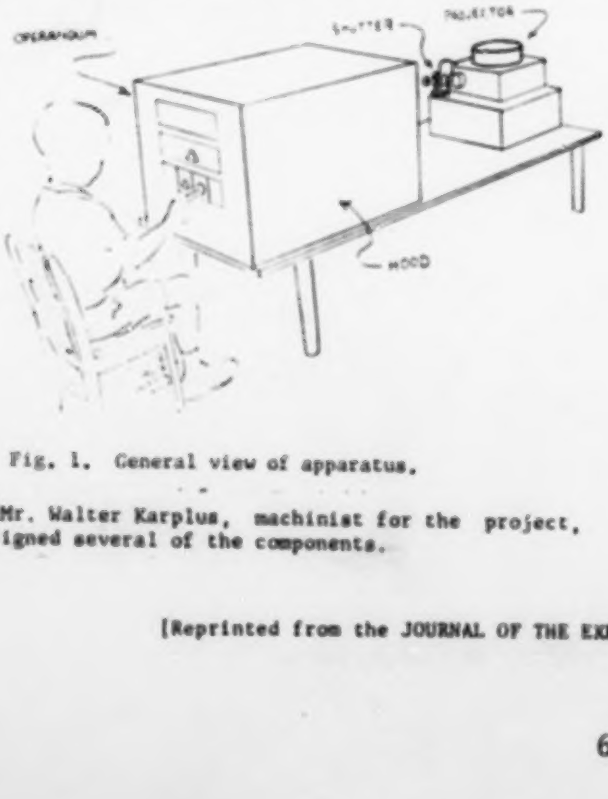

OPERANDUM
SHUTTER
PROJECTOR
HOOD

Fig. 1. General view of apparatus.

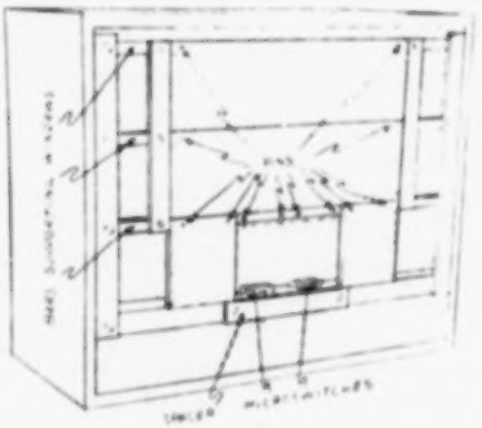

PINS
MICROSWITCHES

Fig. 2. Front view of operandum with faceplate and two center choice-windows removed.

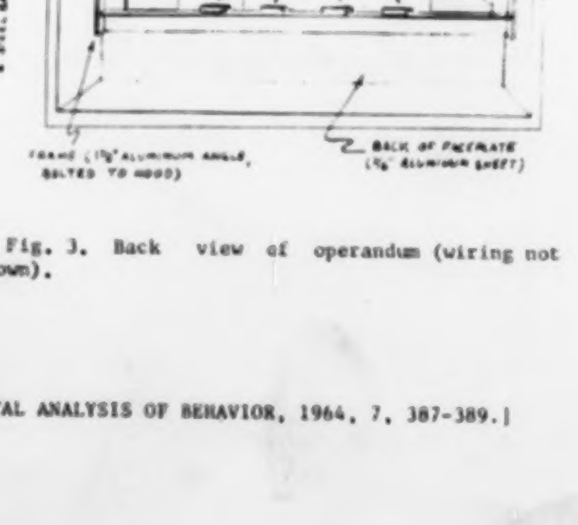

HOOD (½" PLYWOOD) (CUT AWAY)
BRACKETS FOR MICROSWITCHES (1½" ALUMINUM ANGLE, CUT TO FIT AROUND BARS)
BRACKET FOR PHOTOCELLS (1½" ALUMINUM ANGLE)
MICROSWITCHES
¼" STEEL BARS SUPPORTING WINDOWS
FRAME (1½" ALUMINUM ANGLE, BOLTED TO HOOD)
BACK OF FACEPLATE (1/16" ALUMINUM SHEET)

Fig. 3. Back view of operandum (wiring not shown).

---

[1] Mr. Walter Karplus, machinist for the project, designed several of the components.

[Reprinted from the JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR, 1964, 7, 387-389.]

the switch housing limits the travel of the window. The spring in the microswitch returns the window to its original position when pressure is removed. Depending upon the size of the window, it may be necessary to locate more than one microswitch along its lower edge and to wire them in parallel. The apparatus shown has two large windows, each with two microswitches, and four small windows, each with one. The windows are held in place by a faceplate, which prevents them from falling off the pins, but does not exert pressure against them. Dotted lines in Fig. 2 indicate openings cut in the faceplate.

Figure 3 shows the placement of the photocells. Wiring for the switches and photocells is carried to a cable connector located in the hood.

Figure 4 shows the lay-out used in making 35mm. transparencies for use in this apparatus. Each set of stimuli, together with one or more coding spots which actuate the photocells is photographed on a single slide. One can think of the lay-out as being superimposed on the front of the operandum: dotted lines indicate the windows. The dimensions shown may be scaled up and down as necessary to photograph stimuli of different sizes. In inches, they correspond to the dimensions of the projected image at the plane of the windows.

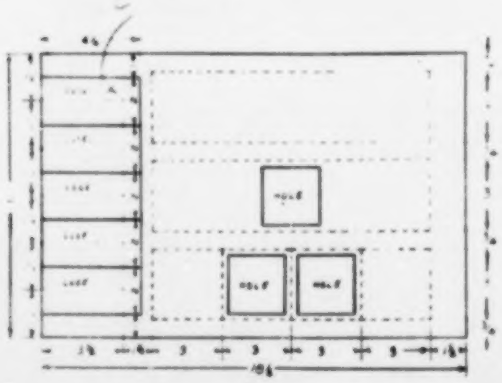


Fig. 4. Photography lay-out.

Figure 5 shows the photography set-up. Following the dimensions of the lay-out, a black cardboard mask is constructed, with holes beneath which the stimuli-to-be-photographed may be placed. (A two-choice matching-to-sample set-up is shown.) White coding-tabs are hinged to the left side of this mask. In production the stimuli are laid on the easel, the mask is placed over them, the appropriate coding-tab or tabs are folded so that they lie over the mask, and these materials are held flat by a sheet of plate glass while the picture is taken. (All surfaces should be painted flat black to avoid reflections in the glass.)

A good deal of care must be taken in photographing and mounting the slides to make sure that the positioning is accurate when the slide is projected. The outside dimensions of the lay-out in Fig. 4 correspond to the dimensions of the standard 35mm. "frame." A good procedure is to outline this rectangle on the mask with white tape and to focus the camera (using a piece of ground glass at the plane of the film) so that the frame just fits the aperture of the camera. On the processed film each

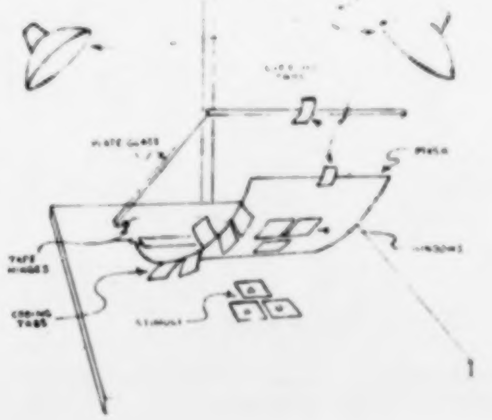


Fig. 5. Photography set-up.

frame will then be outlined in white and this may be used to center the film in the cardboard mount. Ordinary commercial mounting is not sufficiently accurate. Some commercial processors will hand-mount the slides according to specifications for an additional fee.

To achieve consistent results when preparing a large number of slides, we have found it necessary to buy a large stock of color film all with the same emulsion number and to store it under refrigeration. The exposed films are sent to a large regional laboratory for development. Films developed by local quick-service vary considerably from one batch to another. The films are returned uncut and hand-mounted locally. If these precautions are taken, the positioning of the slides is accurate enough for most purposes.

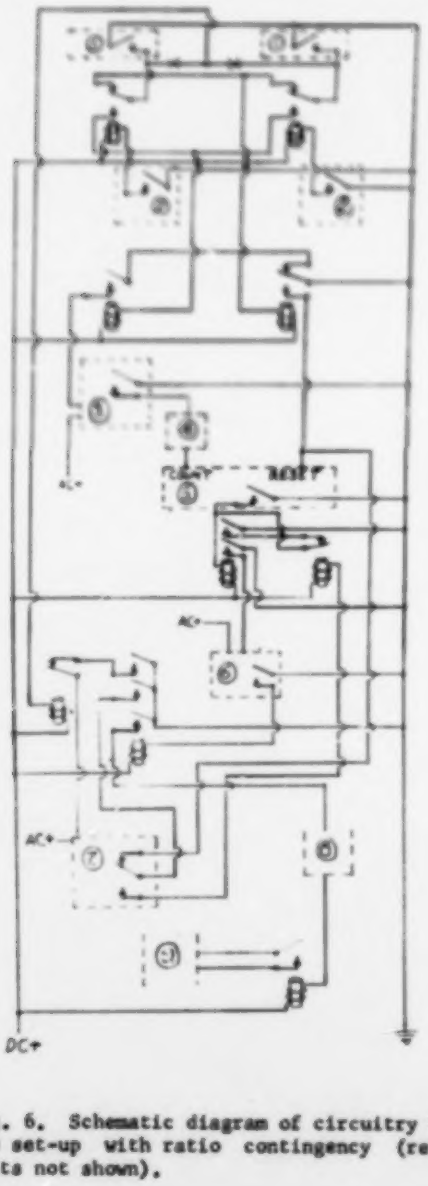


Fig. 6. Schematic diagram of circuitry for two-choice set-up with ratio contingency (recording circuits not shown).

The *Kodak Carousel* Projector is quite reliable. It also reverses automatically, and if one wishes to do so, the apparatus can be wired so that it reverses upon command from a photocell, i.e., when it reaches a given slide.

A schematic wiring-diagram for the two-choice, matching-to-sample task is shown in Fig. 6. This circuit requires the subject to respond to the correct choice a given number of times before the projector will operate--a contingency which was found to improve performance considerably (Hively, 1963). Recording circuitry is not shown in the diagram.

## REFERENCES

Gollub, L. R. A simple pulse shaper. *J. exp. Anal. Behav.*, 1958, *1*, 122.

Hively, W. Programming stimuli in matching to sample. *J. exp. Anal. Behav.*, 1962, *5*, 279-298.

Hively, W. Parametric experiments with a matching to sample program. Harvard University Teaching Machine Project, Final Report, 1964.

PRECEDING PAGE BLANK-NOT FILMED

# PROGRAMMING STIMULI IN MATCHING TO SAMPLE[1]

**Wells Hively**

In these investigations, a "teaching machine" was used to train pre-school and first-grade children in a series of progressively difficult discrimination tasks, leading up to matching to sample. Such training was much more efficient than training in the final discrimination alone. The errors the subjects made were found to be a functon both of the differences between consecutive discriminations (the "size of the steps" in the program) and the length of training on each discrimination. Theoretical and practical implications of these findings are discussed.

"Teaching machines" were originally conceived as devices for applying powerful principles and techniques, derived from the study of learning in animals, to human instruction (Skinner, 1954); and they are ordinarily discussed in those terms (*cf.* Holland, 1960). Less has been said, perhaps because it has been taken for granted, about the use of teaching machines as instruments for studying human learning. Besides being used to apply known principles to practical tasks, teaching machines may be used as other pieces of operant-conditioning apparatus are used: to control the conditions under which learning occurs in order to study it better. Such research is not simply concerned with assessing the relative effectiveness of alternative programs or techniques. It is also concerned with showing how the results of one experiment are related to those of another.

The present research was the outgrowth of an effort to study concept formation in children while retaining contact with research on discrimination learning in animals. It made use of a type of teaching machine originally conceived by B. F. Skinner. The behavior studied was matching to sample.

## EXPLORATORY EXPERIMENTS

### *Apparatus*

Figure 1 shows the apparatus. A slide projector inside the box projected stimuli onto the windows in the front panel. The top section of the panel was not used in these experiments. The center section consisted of a single, long window in which the sample stimulus appeared. The bottom section was di-

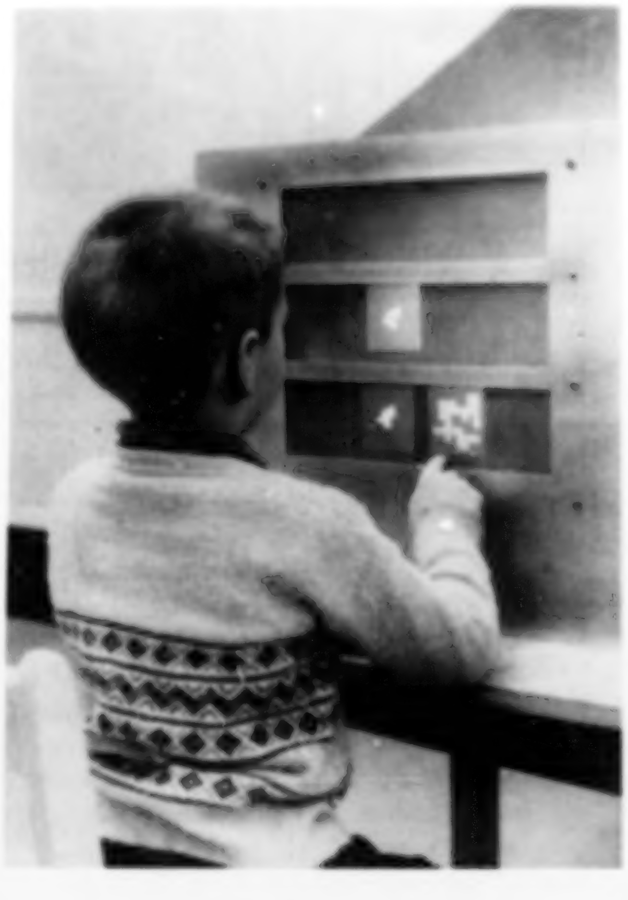


Fig. 1. Apparatus.

---

[1] This article is based on a thesis presented to the Committee on the Ph.D. in Education, Harvard University, April, 1961. The research was supported by grants to the Harvard Department of Psychology from the Carnegie Corporation and the U. S. Office of Education. It was made possible through the Cooperation of the Harvard Graduate School of Education's School University Program for Research and Development, the Lexington, Massachusetts, Public Schools, and the Lesley Ellis School, Cambridge, Massachusetts. Special thanks are due to Mrs. Anne Connors and Mrs. Helen Popp, who acted as experimenters.

[Reprinted from JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR, Vol. 5, No. 3, July 1962.]

vided into four small windows, and the choice stimuli appeared in the middle two of these. The subject's task was to select the one of the two lower stimuli which matched the one above. He indicated his choice by pressing the window on which the stimulus appeared. This action closed a switch, and if the choice was correct, the projector automatically changed slides. (The apparatus detected the position of the correct choice photoelectrically.) If the choice was incorrect, nothing happened, and the subject could respond again immediately.

An approximately 270-g force, applied to the center of a choice window, was required to close the switch, which made a clearly audible "click." The control circuit required that a switch be held closed for 0.15 sec in order to record a correct choice. During this time, an incorrect response broke the circuit and reset the timer. Thus, subjects were prevented from operating the machine by simply pressing both choices at once. Changing slides required 0.75 sec, during which time the projection beam was broken by a shutter which remained closed until all window switches returned to the normal position. Responses and slide changes were recorded on a polygraph.

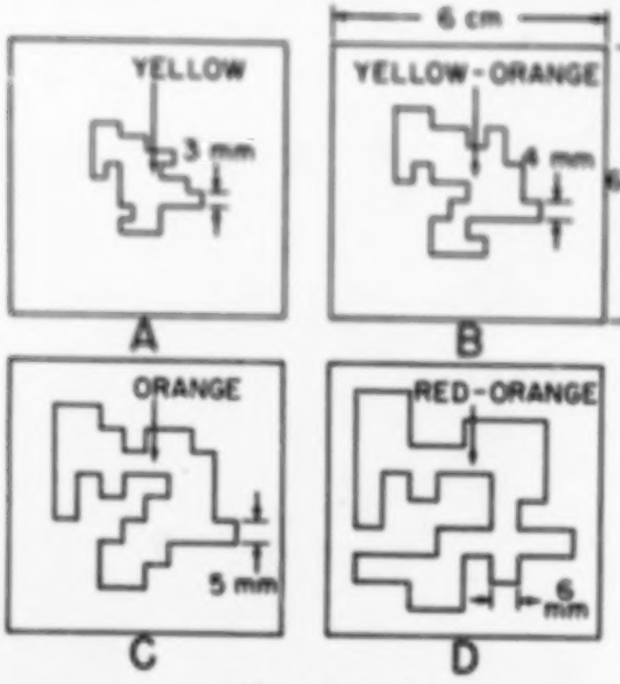


Fig. 2. Stimuli.

Figure 2 shows the stimuli, which differed systematically in shape, size, and color. They were constructed after a method described by White (1957): 32 cells were filled in an 8 by 8 matrix, and then 4 cells at a time were moved to new locations, starting from Form A and ending with Form D. These forms were plotted in successively larger matrices, transferred to *Color-Aids* papers of neighboring hues, mounted on uniform grey backgrounds, and photographed on *Kodachrome* film. They were projected "life" size.

*Procedure*

Subjects were brought to an experimental room one at a time by an experimenter who talked with them freely but did not discuss the experiment. In the room, she gave the following instructions:

(1) "This is a machine for you to play with." The experimenter turned on the projector, showing the first slide.

(2) "You can make it go like this." The experimenter operated the apparatus deliberately and correctly four times.

(3) "You do it." If the subject did not respond or stopped responding, the experimenter repeated instruction No. (2).

(4) After the subject had operated the apparatus four times, the experimenter said, "Good, see if you can push the *right one every* time," and retired to a chair about 8 ft behind the subject.

(5) Remarks the subject made during the session were ignored insofar as possible, or answered by, "See if you can push the right one every time."

(6) If the subject asked to leave, the experimenter said, "Wouldn't you like to do the rest?" If the subject insisted, he was allowed to go. At that time, or at the end of the session, the experimenter said, "All finished? Good, perhaps you can come back again another time."

Subjects were members of the first grade in the Franklin School, Lexington, Massachusetts. The 94 children were listed in order of age, from the youngest (5-8) to the oldest (6-10), and the list was divided into six equal groups. One child was chosen at random from each of these age groups to make up each experimental group of six subjects.

*Experiment I.* This experiment presented a standard matching-to-sample problem in which each of the four stimuli appeared as a correct choice, in contrast to each of the others and in each of the choice windows, in

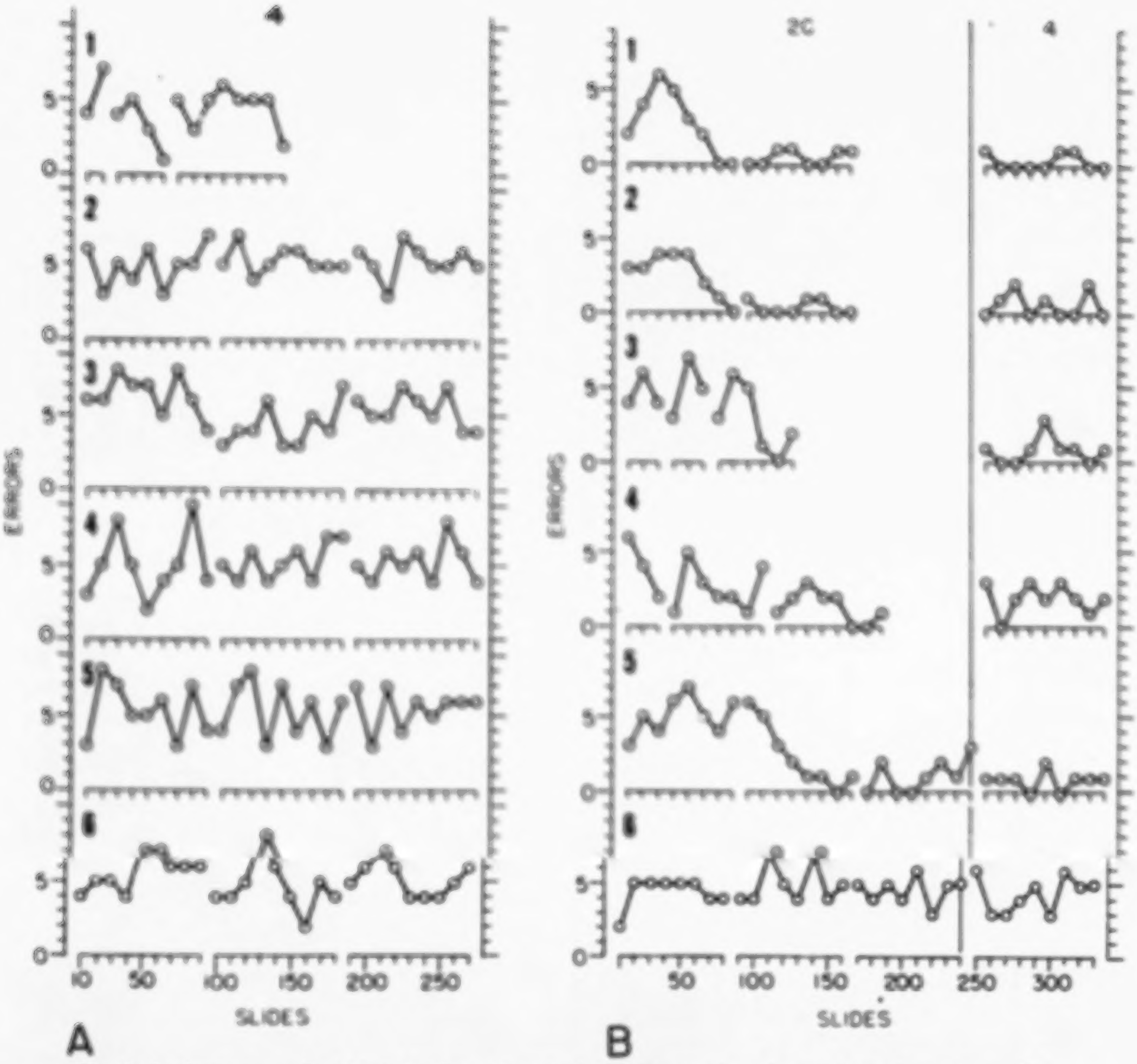


Fig. 3. Performances of subjects in Experiments I and II: A, Experiment I; B, Experiment II.

a "random" sequence. There were therefore 24 possible stimulus combinations (identified by two letters indicating the choice stimuli and their position, followed by a hyphen and a letter indicating which appeared as the sample): AB-A, AB-B, BA-A, BA-B; AC-A, AC-C, CA-A, CA-C; AD-A, AD-D, DA-A, DA-D, BC-B, BC-C, CB-B, CB-C; BD-B, BD-D, DB-B, DB-D; CD-C, CD-D, DC-C, DC-D. Each of these items was photographed four times. A Gellerman Series (Gellerman, 1933) was used to determine a sequence of right and left correct-choice positions, and slides were selected at random to fill each position. The resulting series of 96 slides was labeled *Series 4*, because it made use of each of the four different stimuli.

Six subjects were given Series 4, and in two subsequent sessions (on separate days) they repeated it. Figure 3A shows the number of errors made by each subject in consecutive groups of 10 slides. The last six slides in the series are not shown. Only errors made on the *first* response to each slide were counted; therefore, a subject responding at random to 10 slides would make 5 errors on the basis of chance. Points within a single session are connected; gaps in the records indicate the beginning of the next session. Subject 1 did not complete the program during any session.

No subject learned this discrimination within the three sessions, but all responded systematically in various ways. The most common were alternation and preservation with

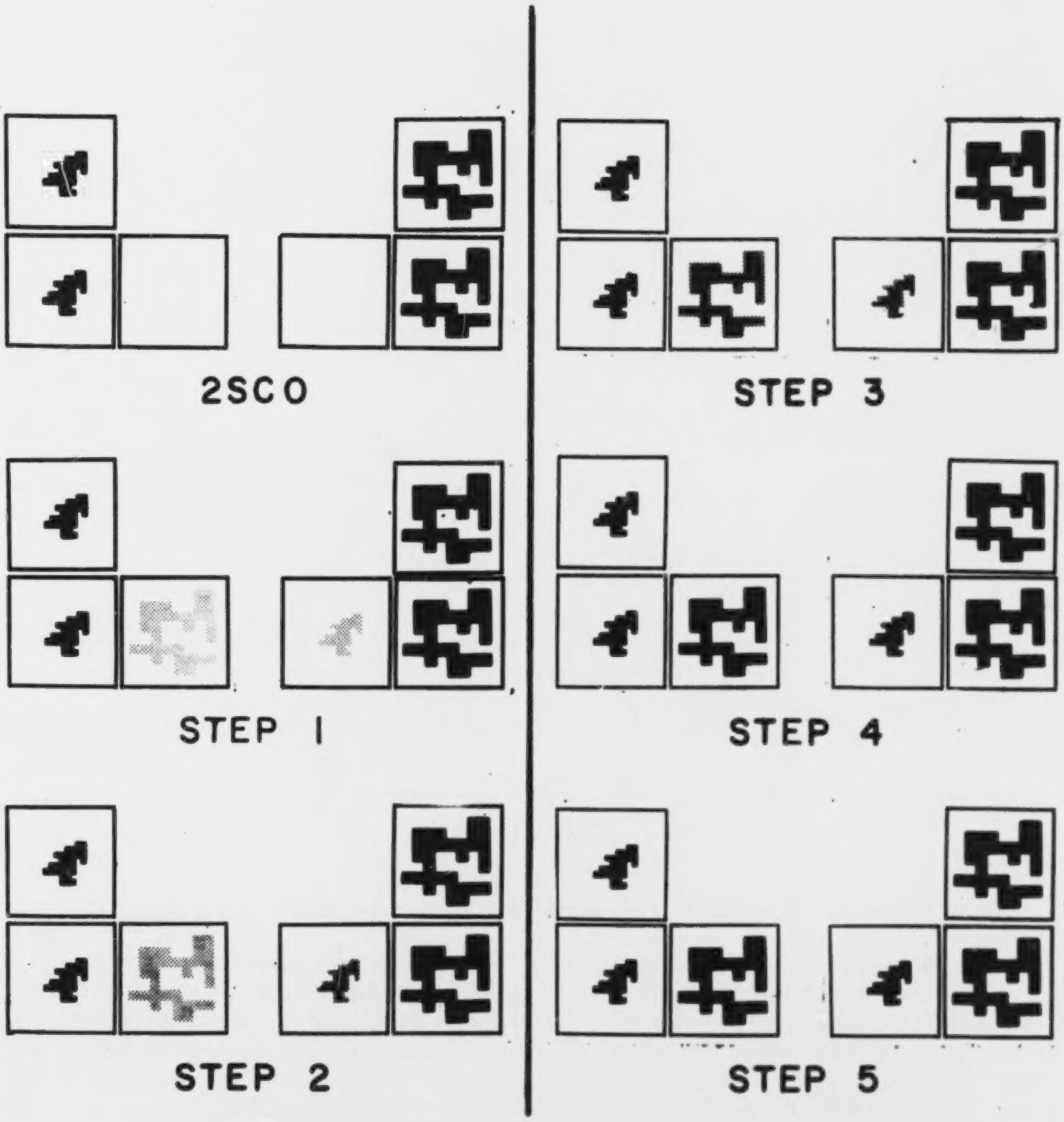


Fig. 4. Items from Series 2SC0, Transition Series 1, and Series 2SC.

respect to the choice windows. However, subjects also favored one window or the other; followed the larger or smaller of the stimuli; pressed both windows at once; or played on the windows as if they were piano keys, switching frequently from one pattern of responding to another during a session.

I thought that subjects might respond to the matching relationship if they began with an easy discrimination and approached the matching task through a series of "successive approximations," a procedure suggested by theoretical discussions of teaching-machine programming (Skinner, 1954; Holland, 1960).

The next experiments were attempts to develop such a sequence of progressively difficult discriminations leading up to the matching-to-sample task required in Experiment I.

*Experiment II.* Series 2C was designed to present an easier task than Series 4. It used only stimuli *A* and *D* (Fig. 2) Stimulus *A* always appeared in the left-hand choice window, and *D* in the right; only the sample changed. The two items used in this series are shown in Fig. 5, where they are labeled 2C(2). The items from Series 2C also are part of Series 2, which will be described later. Other items shown in Fig. 4 and 5 are from experiments which follow. Each was photographed 40 times, and the Gellerman series was used to determine the sequence in which the 80 slides appeared. The number "2," identifying this series of slides, indicates that two stimuli were used, and the letter "C" indicates that the position of the stimuli in the choice windows was constant.

Six subjects were given Series 2C, and repeated it in subsequent sessions on separate days. After they had reached a criterion of 10 per cent error over the entire series, or had completed three sessions, they were given Series 4. Figure 3B shows their records. Subject 4 did not complete the series during Session 1, and Subject 3 did not complete it during any session.

Five subjects learned the task required by Series 2C; and when they were switched to Series 4, their records were similar to their records at the end of Series 2C. A comparison of Fig. 3A and 3B shows that the task required by Series 2C was in fact easier than that required by Series 4, and that pre-training on Series 2C facilitated the discrimination required in Series 4.

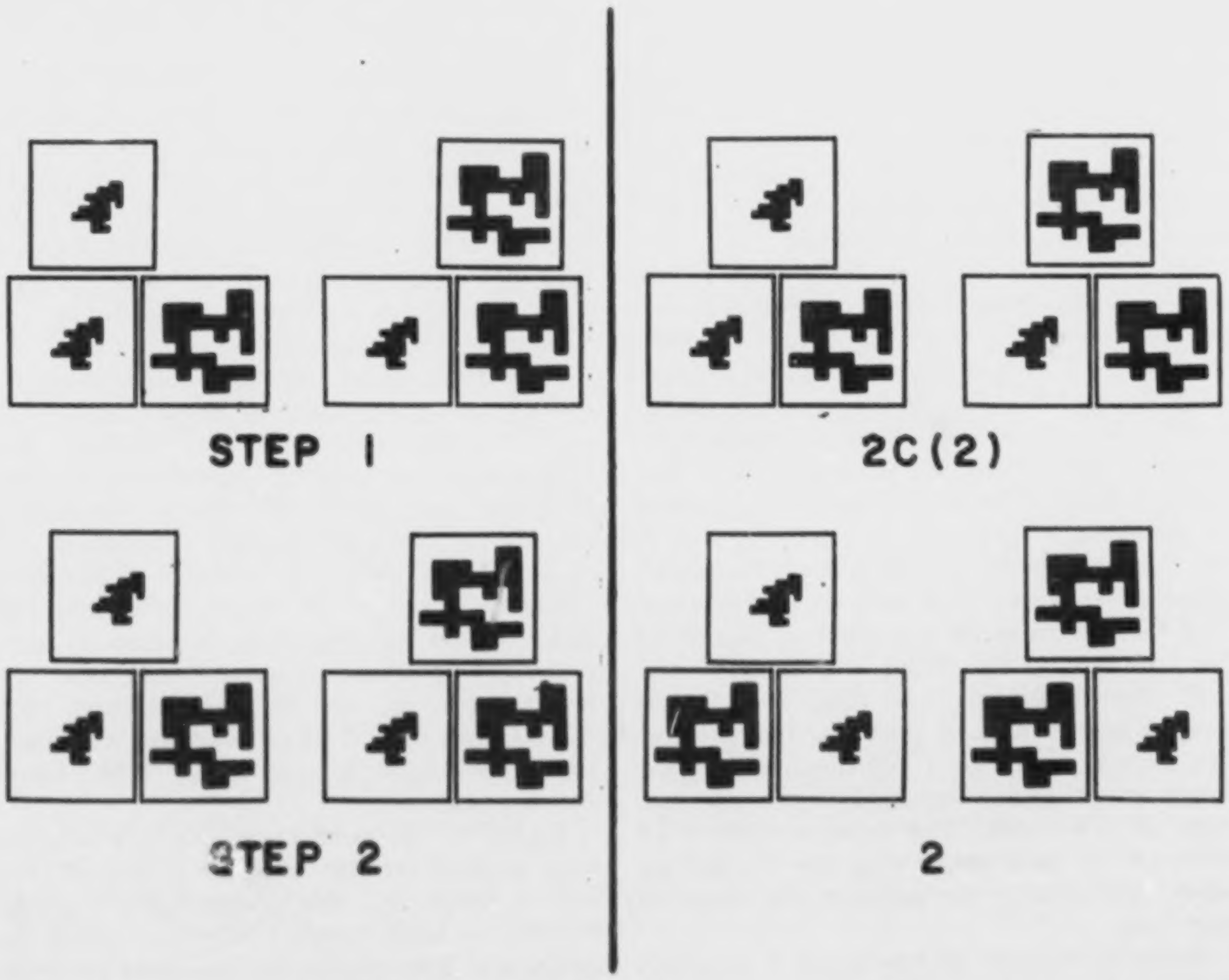


Fig. 5. Items from Transition Series 2, Series 2C, and Series 2.

*Experiment III.* Series 2SC0 and 2SC were designed to precede Series 2C. In Series 2SC the sample was located directly above the correct choice, rather than midway between the two. The letter "S," identifying the series, refers to this spatial relationship between the sample and the correct choice. Figure 4 shows the two items used in this series. The series contained 70 slides arranged in sequence using the Gellerman series.

Series 2SC0 was identical to Series 2SC, except that no incorrect choice stimulus appeared. The figure "0" refers to the absence of an alternative. The two items used in this series are also shown in Fig. 4.

Series 2SC0, 2SC, 2C, and 4 were presented in that order, all in one session. The apparatus shut off at the end of each series, and the experimenter said, "Good, now if you'll wait just a minute, I have some more for you to do." She then changed slide magazines and presented the next series.

Figure 6A shows the records of the six subjects who worked on this "program." All subjects responded correctly most of the time. The most errors occurred at the transitions from 2SC0 to 2SC and from 2C to 4, although there was considerable individual variation.

*Experiment IV.* Additional series were added to supply transitions from 2SC0 to 2SC, from 2SC to 2C, and from 2C to 4.

In the transition from 2SC0 to 2SC, the incorrect choice was introduced through four steps of increasing intensity. Figure 4 shows these steps. The two items were photographed with *Bourges Overlay Sheets* covering the incorrect choice-stimulus. These are transparent acetate sheets which are printed with a fine-dot pattern; the density of the dots provided 10 per cent, 30 per cent, 50 per cent, and 70 per cent grey "screens." Each item was photographed six times with each of the overlays. The Gellerman series was used to determine the sequence of the slides within each step.

In the transition from 2SC to 2C, the sample moved from a position directly over the correct choice to a position midway between the choices. Figure 5 shows the two steps in this series. The series contained 16 slides, 8 at each step, with the Gellerman series determining the sequence of slides in each step.

Series 2C was followed by Series 2, in which the stimuli now appeared at random in either choice window. Figure 5 shows the four items used in Series 2. Each was photographed four times, and the 16 slides were assigned at random to positions in the Gellerman series.

Series 2 was followed by a transition series in which Stimuli *B* and *C* were introduced, but only *B* appeared in contrast to *D* and *C* in contrast to *A*. This series used eight items:

| | | | |
|---|---|---|---|
| BD-B, | BD-D, | DB-B, | DB-D |
| AC-A, | AC-C, | CA-A, | CA-C |

Each was photographed twice, and the 16 slides were randomly assigned to the Gellerman series.

The series were presented in order: 2SC0 (30 slides), Transition (24 slides), 2SC (40 slides), Transition (16 slides), 2C (40 slides), 2 (16 slides), Transition (16 slides), and 4 (94 slides). The first 130 slides (through the first 20 slides in Series 2C) were presented without interruption. The apparatus then shut off, and the experimenter said, "Good, now if you'll wait just a minute, I have some more for you to do." Then she presented the last 146 slides.

Figure 6B shows the records of the six subjects who worked on this program. Adding the transition series reduced the errors between Series 2SC0 and Series 2C and caused them to be distributed more widely. The total number of errors was reduced through Series 2C, but considerably more errors were made on Series 4 than were made by the subjects in Experiment III.

*Experiment V.* A number of slides were removed from the ends of Series 2SC0, 2SC, and 2C; but, otherwise, the material was identical to that used in Experiment IV. The program was: 2SC0 (20 slides), Transition (24 slides), 2SC (20 slides), Transition (16 slides), 2C (20 slides), 2 (16 slides), Transition (16 slides), and 4 (94 slides). The first 132 slides (up to the beginning of Series 4) were presented without interruption. The apparatus shut off, and the experimenter said, "Good, now then if you'll wait just a minute, I have some more for you to do." Then Series 4 was presented.

Figure 6C shows the records of six subjects who worked on this program. Through the end of Series 2C, this group made approximately the same number of errors as did the subjects in Experiment IV, but their performances on Series 4 were strikingly superior.

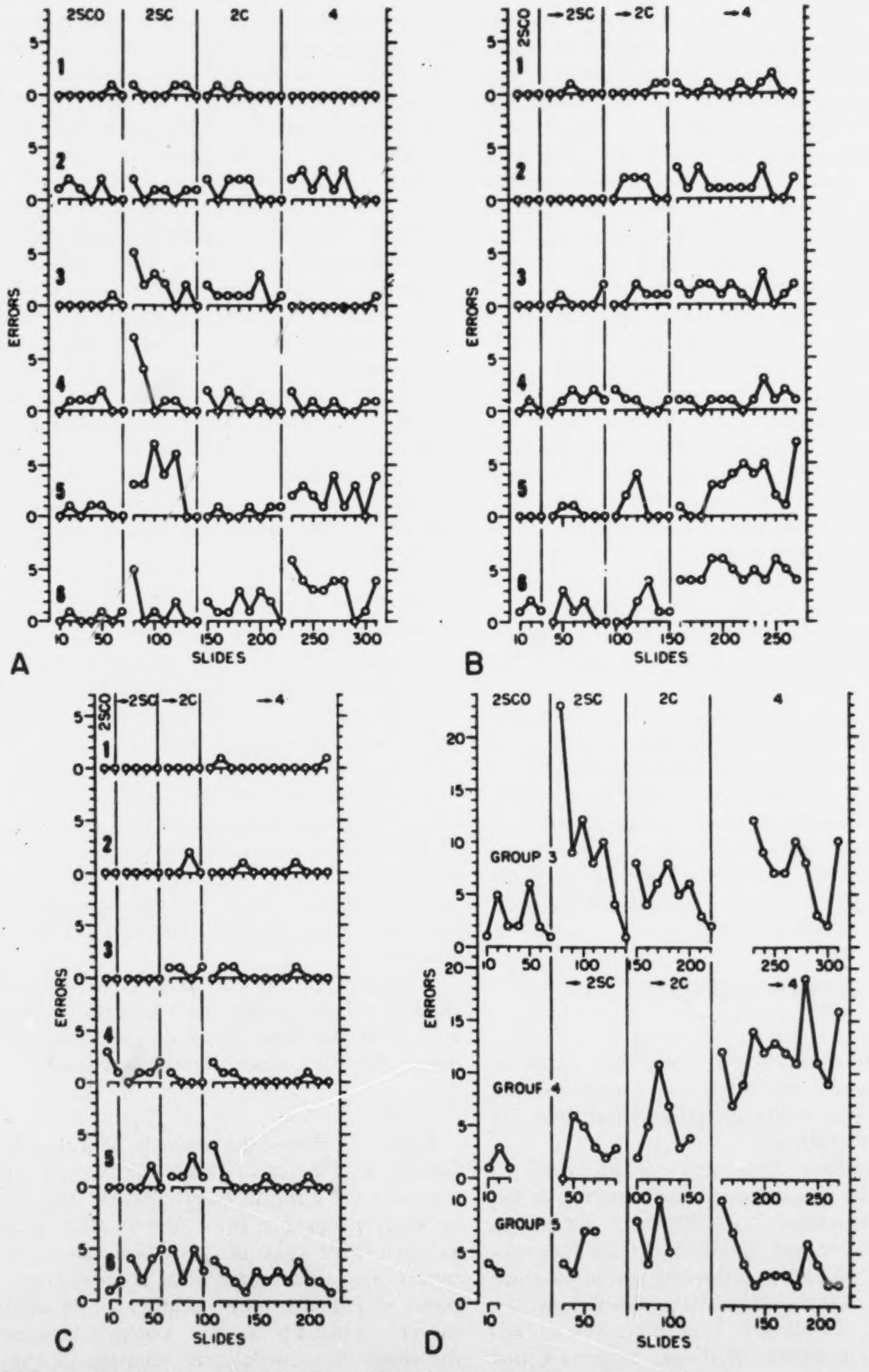


Fig. 6. Performances of subjects in Experiments III, IV, and V: A, Experiment III; B, Experiment IV; C, Experiment V; D, Group Performances.

The performances of subjects in Experiments III, IV, and V are summarized in Fig. 6D, which shows the total number of errors made by each group on comparable sections of their programs. The record of Group 3 is what one would expect from a set of transfer experiments: errors increase at the points of transition between series, and decrease throughout each series. The records of Groups 4 and 5 are considerably more irregular. An overall increase in errors is evident in the record of Group 4.

Putting the children through a program of progressively difficult discriminations clearly increased the efficiency with which they learned the matching task. Apparently, the smaller the "steps," *i.e.*, the less difference between consecutive discriminations, the more efficient the program. But the effect of shortening sections of the program in Experiment V was quite suprising: Why should fewer errors occur in a short program than in a long one, if the two are otherwise identical? If length is a variable, did Group 5 perform better than Group 3 because the program had smaller steps, or because the steps were shorter, or both?

Two artifacts were possible in the exploratory experiments: One, subjects who learned the task might have told their friends how to do it; and, two, the experimenter might have become more skilled at instructing the subjects during the experiments. A control experiment was designed to replicate the exploratory experiments and to study the effect of program length somewhat further.

## CONTROL EXPERIMENT

### *Programs*

The *Criterion Program* was the same as Series 2 (Fig. 5): a random matching-to-sample task with two stimuli, rather than four. It contained 120 slides.

The *Standard Program* was identical to that used in Exploratory Experiment V, except that it ended with Series 2, to which four slides were added. The Standard Program thus consisted of the following series, totaling 120 slides: 2SC0 (20 slides), Transition (24 slides), 2SC (20 slides), Transition (16 slides), 2C (20 slides), and 2 (20 slides). Figures 4 and 5 show the items from each step in this program in order.

The *Expanded Program* was the same as the Standard Program, except that each series was approximately twice as long. It totaled 243 slides arranged in order: 2SC0 (48 slides), Transition (48 slides, 12 at each step), 2SC (48 slides), Transition (32 slides, 16 in step 1, 15 in step 2), 2C (48 slides), and 2 (20 slides).

The *Reduced Program* was the same as the Standard Program, except that each series was approximately half as long. It totaled 84 slides arranged in order: 2SC0 (12 slides), Transition (16 slides, 4 at each step), 2SC (12 slides), Transition (12 slides, 6 at each step), 2C (12 slides), and 2 (20 slides).

The *Abridged Program* was the same as the Standard Program except that all transition series were deleted and replaced by an equal number of slides from the series which followed. It therefore totaled 120 items: 2SC0 (20 slides), 2SC (44 slides), 2C (36 slides), and 2 (20 slides).

### *Procedure*

The subjects were children in the Lesley Ellis School, Cambridge, Massachusetts. They were drawn from a kindergarten class, two "transition" classes, and two first-grade classes; and their ages ranged from 4-6 to 7-1. Two subjects were chosen at random from the kindergarten and one from each of the other classes to make up each experimental group of six. Table 1 shows the format of the experiment. Groups were run in the approximate order in which they are presented in the table, and a day or more elapsed between sessions.

The instructions and procedure were identical for all groups (except where noted below), and were the same as those described previously. Another experimenter was used.

### *Results*

Figure 7 shows the records of subjects in Group S. The errors each individual made are plotted cumulatively against the slides in each program. Intervals on the vertical axis are twice as large as intervals on the horizontal axis. The height of a curve at any point shows the total number of errors the subject made up to that point. Changes in the slope of a curve show changes in the accuracy with which the subject responded. Chance performance would yield a slope of

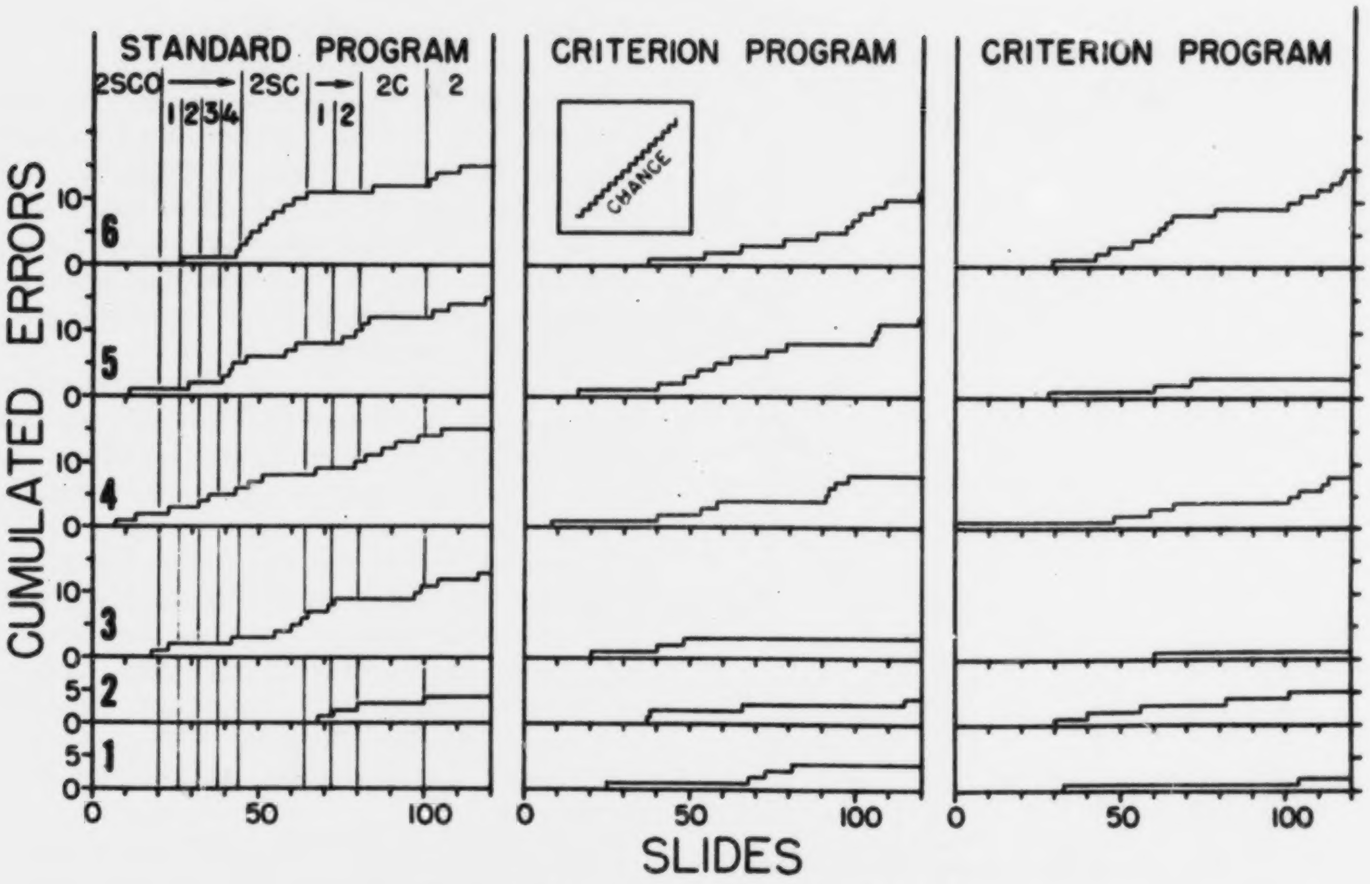


Fig. 7. Performances of subjects in Group S.

45 degrees. (See Inset.) During periods when a subject was making no errors, his cumulated error curve is a horizontal line.

The number of errors made by subjects in Group S was not significantly different from that made by subjects working on the identical portion of the program in Exploratory Experiment V. The effectiveness of the Standard Program therefore cannot be attributed to communication between the children or to the behavior of the experimenter. [All comparisons, except where otherwise noted, are by the Mann-Whitney Test, using the 5% level of significance, *cf.* Mostellar and Bush (1954). For cases in which the results were predictable on the basis of the exploratory experiments, a "one-tailed" test was used.]

All subjects in Group C performed at a chance level throughout the first two sessions. The fewest errors any of the subjects in Group C made during either session was 60, whereas no subject in Group S made more than 15. This confirmed the major finding of the exploratory experiments: approaching the criterion task through a sequence of progressively difficult discriminations was much more efficient than training in the criterion task alone.

After two sessions on the Criterion Program, Group C was switched to the Standard Program. At the beginning of this session, the experimenter said, "Let me show you how to do it," and then repeated the usual instructions.

Figure 8 shows the records of Group C for this session. (The record of Subject 6 is set back to the base line at slide 100.) The effect of their previous work on the Criterion Program is clear: transfer to the Standard Program was negative. Only two subjects in Group C showed evidence of coming under control.

Figure 9 shows the records of subjects in Group A. (The record of Subject 6 in Session 2 is set back to the base line at slide 79.) Subject 4 did not complete the program during Session 1, and Subject 5 declined to return for Session 2. The number of errors these subjects made on the Abridged Program was significantly larger than that made by Group S on the Standard Program. They also made a significantly larger number of errors on the Criterion Program. Deleting the transition series increased the errors both in training and on the criterion task. Making the

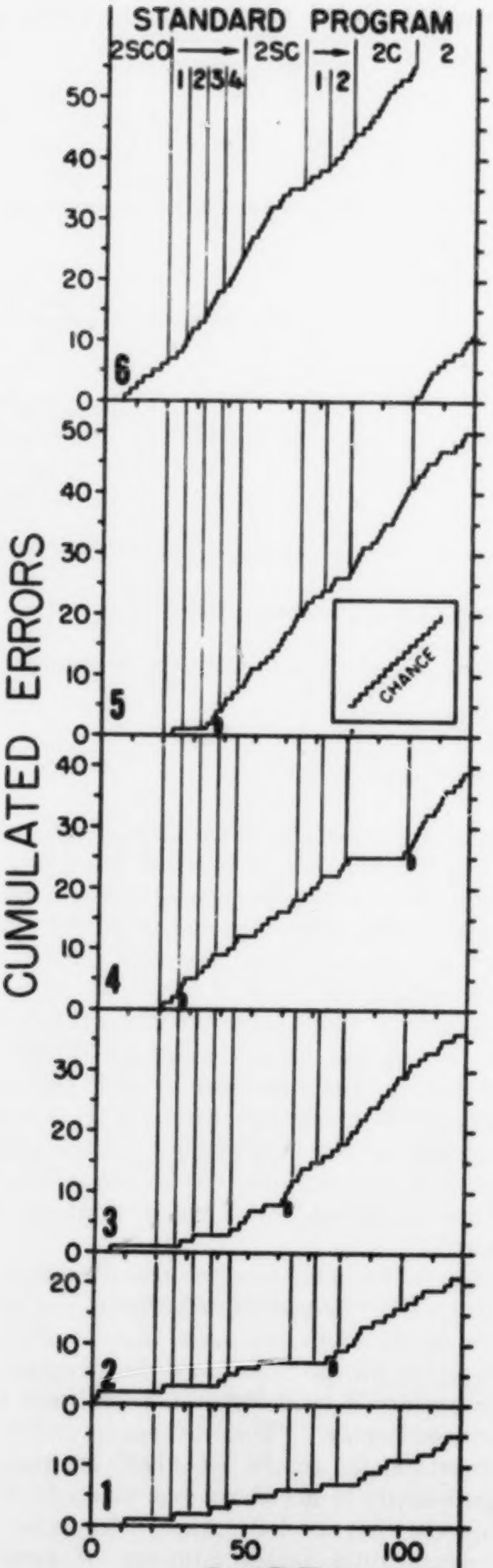


Fig. 8. Performances of subjects in Group C, on Standard Program.

training programs equal in length amplified rather than decreased the effect.

Figure 10 shows the performances of subjects in Group E. The number of errors these subjects made in working through the Expanded Program was significantly larger than the number of errors the subjects in Group *S* made in working through the Standard Program. This confirmed the trend found in exploratory experiments.

One explanation of this difference might be that once a subject has learned the task required by any given series of slides the errors occur at random. A longer program would therefore allow more random errors to occur. However, the number of errors the subjects in Group E made on the Expanded Program (243 slides) was also significantly larger than the number of errors the subjects in Group S made on the Standard Program and the Criterion Program taken together (240 slides). Therefore, that explanation must be rejected.

Some of the errors Group E made might have been due to fatigue. The average length of Session 1 for Group S was 5 min, whereas it was 11 min for Group E. This explanation is ruled out by comparing the number of errors made by the two groups in Session 2, when both were fresh. The number of errors Group E made was significantly larger in this session as well.

The effect does not seem to be due simply to repetition of a given series. The number of errors Group E made on the Criterion Program in Session 2 was also significantly larger than the number of errors Group S made on the Criterion Program in Session 3, although at the beginning of Session 3, Group S had already worked through 140 slides from Series 2, whereas Group E had worked through only 20.

It still seemed possible that the lengths of the sessions might be contributing to the effect. Group S2 was run as a control for this ill-defined possibility. This group was given the Standard Program and two repetitions of the Criterion Program (totaling 360 slides) all during a single session which averaged 17 min in length. Figure 11 shows the performances of these subjects.

When the number of errors Group S2 made on the Standard Program and the first Criterion Program combined was compared

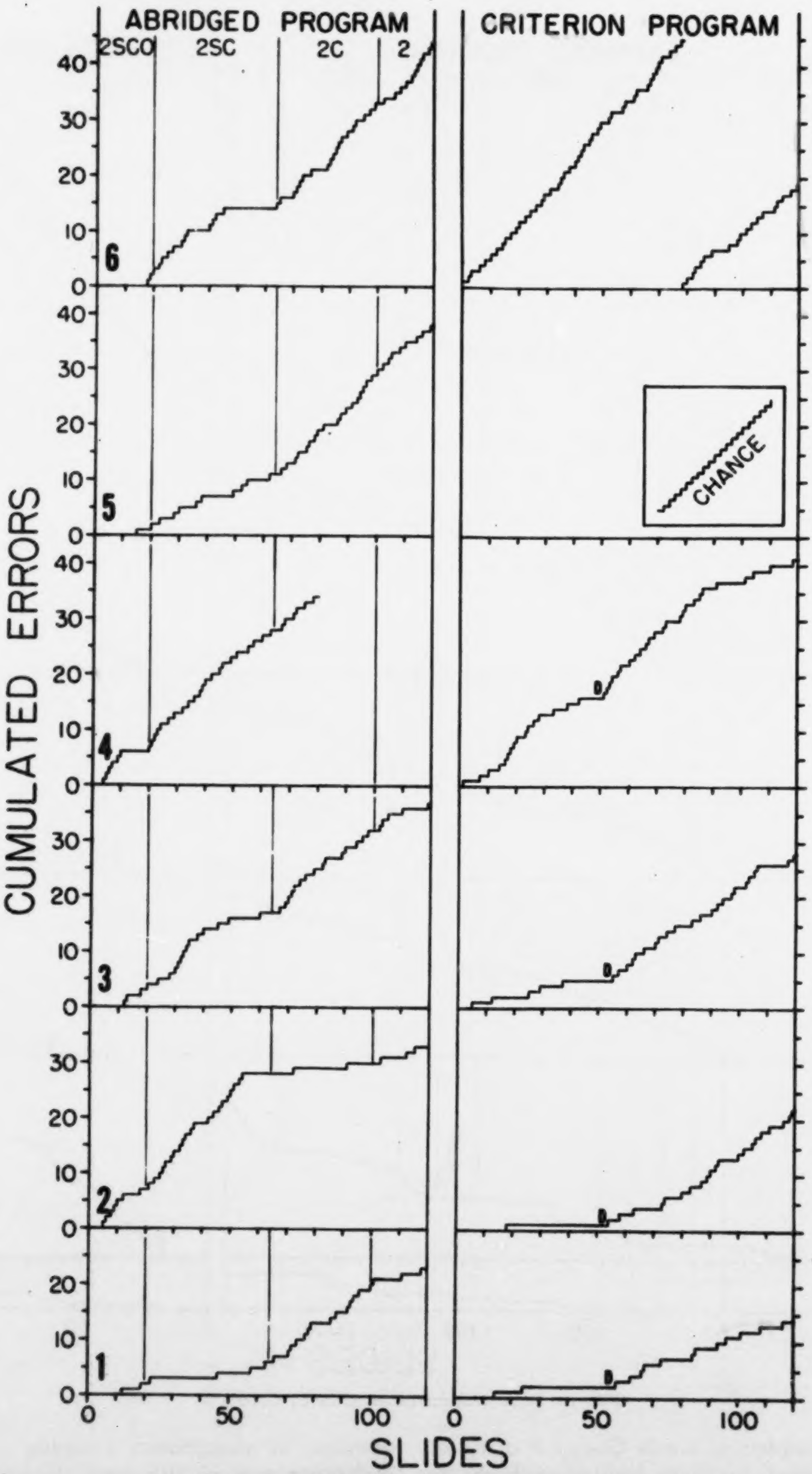


Fig. 9. Performances of subjects in Group A.

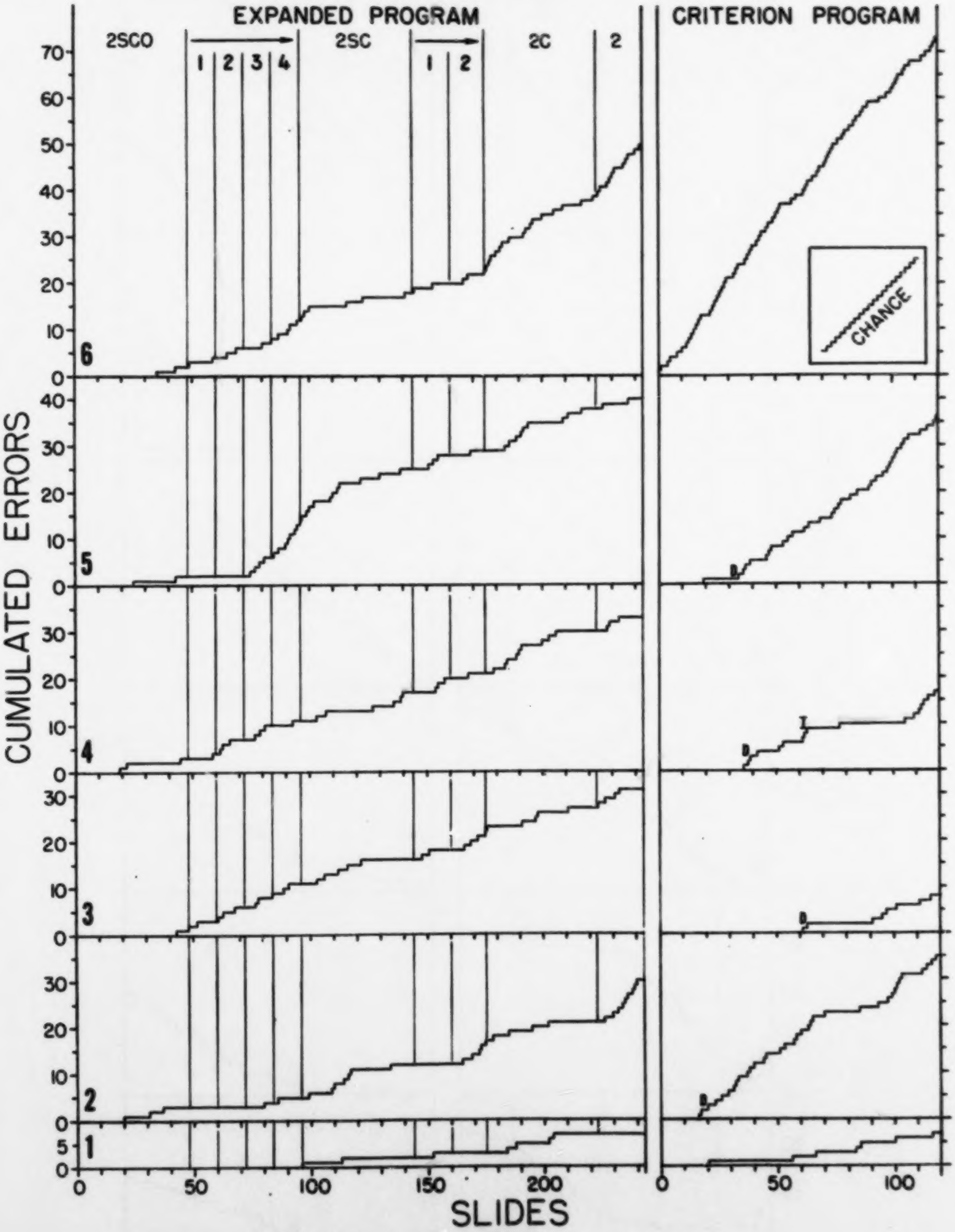


Fig. 10. Performances of subjects in Group E.

with the number of errors Group E made on the Expanded Program (equating both for the total length of the programs and the amount of continuous running time), the difference was in the same direction as between Group S and Group E, but it fell short

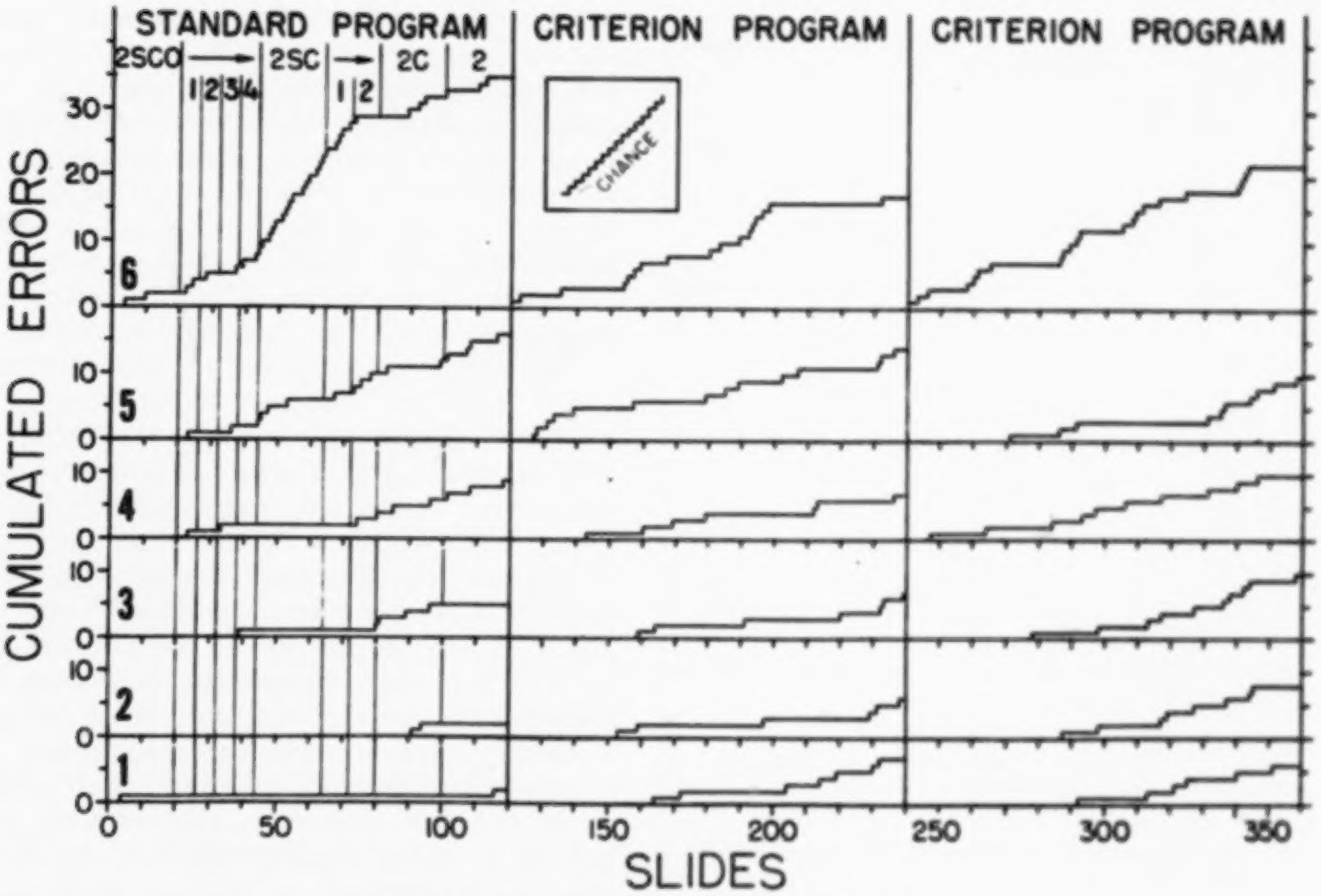


Fig. 11. Performances of subjects in Group S2.

of significance. Comparing the number of errors Group S2 made on its second Criterion Program with the number of errors Group E made on its Criterion Program gave the same result, although in this case fatigue should have given the advantage to Group E.

If presenting all the programs within a single session rather than on separate days had any effect, it would have had to show itself in an increase in the number of errors from program to program in Group S2 compared with Group S. But the errors did not increase through the three programs for either group, and there was no reliable difference between the number of errors made by the two groups on the last program. If massing the trials had any effect upon the number or errors, it was certainly a very small one.

Figure 12 shows the performances of the subjects in Group R. Shortening the program also resulted in an increase in errors. The difference between the number of errors made by Group R on the Reduced Program and that made by Group S on the Standard Program was not significant, but the difference in errors made by the two groups on the Criterion Program was quite large (and significant).

## DISCUSSION

Krechevsky (1932) was one of the first to point out that the behavior of subjects in the "pre-solution period" of discrimination problems is not "random." In the records of individual rats, he found periods of systematic responding with respect to properties of the experimental situation other than the stimuli being manipulated by the experimenter, *e.g.*, position alternation, and perseveration habits. He noticed that the animals often responded systematically in several different ways before coming under control of the relevant stimuli, and he called these periods of systematic responding "hypotheses." Similar behavior was markedly evident in these experiments.

Skinner (1948) and Skinner and Morse (1957) have shown how "incidental" contingencies of reinforcement may straighten systematic responding in a free-operant situation. Periods of systematic responding in a discrimination can be interpreted in a similar way.

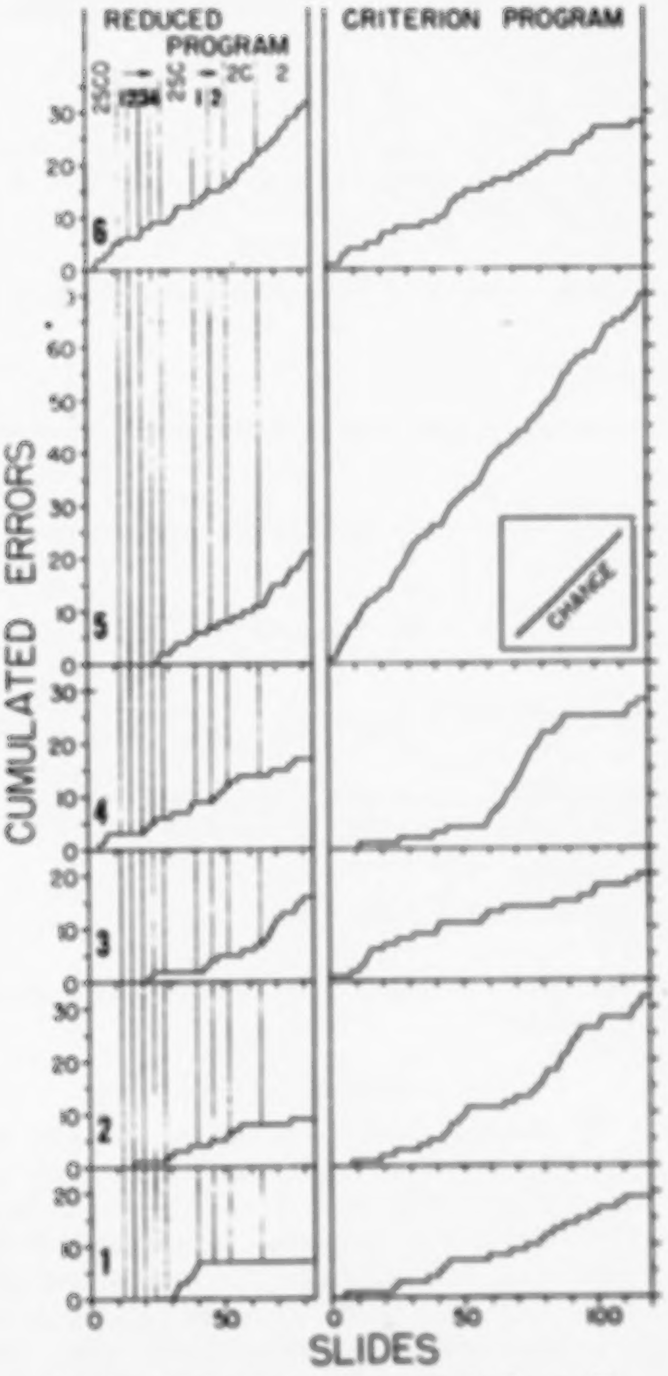


Fig. 12. Performances of subjects in Group R.

The experimenter defines the "correct hypothesis" in a discrimination experiment by arranging to reinforce a given class of responses in the presence of a given class of stimuli. In most discrimination experiments, however, a subject can respond to a variety of stimuli in a number of different ways, some of which may incidentally satisfy the requirements for reinforcement. In these experiments, pressing a choice window was reinforced by operation of the slide changer only when the stimulus occupying that window was identical to the sample stimulus in the window above. However, there was no guarantee that subjects would observe the matching relationship. If a particular response to the correct choice window happened to be under the control of the properties of the specific choice stimulus, or of the spatial location of the window, it, too, was reinforced. Behavior which is thus incidentally strengthened might reasonably be called, in Krechevsky's terms, an "incorrect hypothesis."

The difference between a "correct hypothesis" and an "incorrect hypothesis" is a matter of the probability that the behavior will be reinforced. "Correct hypotheses" are consistently reinforced, and "incorrect hypotheses" received only chance reinforcement. In practice, the problem of minimizing the chances that "incorrect hypotheses" will be accidentally strengthened is complicated by the fact that the "randomness" of a series of correlations between the occurrence of an irrelevant stimulus and the availability of reinforcement depends upon the length of the series. Portions of a series selected with the help of a random number table, for instance, may be nonrandom enough to strengthen "incorrect hypotheses." A short, behaviorally random, sequence of correlations is quite difficult to design.

Furthermore, no matter how carefully one designs a sequence of correlations between the *occurrence* of stimuli and the *availability* of reinforcement, the actual contingencies of reinforcement in a given case depend upon what the subjects observes, which in turn depends upon the individual subject's history. From the experimenter's point of view, it is a matter of chance. Therefore, the chances that "incorrect hypotheses" will be strengthened are a function of the number of stimuli present. The more complex the discrimination problem, the less likely the subjects are to observe the relevant stimuli, and the more likely "incorrect hypotheses" will be incidentally reinforced.

Presenting the subject with a sequence of progressively complex discrimination problems may be thought of as a way of increasing his chances of observing the relevant stimuli in each problem. The increase in efficiency with which subjects in these experiments learned the matching task when it was preceded by such a program is therefore a measure of the degree to which the procedure prevented "incorrect hypotheses" from interfering with the development of the discrimination.

If hypotheses were rejected by a logical process—if, for instance, one correct response "confirmed" a hypothesis, and one incorrect response "rejected" it—the number of errors subjects made in learning a complex discrimination would be larger and more variable than the number of errors subjects made in learning a simple one; but the longer a subject worked on either problem, the closer he would come to the solution. Although none of the subjects in Group C learned the task required by the Criterion Program within two sessions, they might eventually have learned if the experiment had continued long enough. However, the performance of Group C on the Standard Program *reduced* the likelihood that they would respond to the relevant stimuli. Two explanations for this finding are possible.

Table 1

Format of the Control Experiment

| *Group* | *Program* | | |
|---|---|---|---|
| | *Session 1* | *Session 2* | *Session 3* |
| Group S | Standard | Criterion | Criterion |
| Group C | Criterion | Criterion | Standard |
| Group E | Expanded | Criterion | |
| Group R | Reduced | Criterion | |
| Group A | Abridged | Criterion | |
| Group S2 | Standard<br>Criterion,<br>Criterion | | |

Table 2

Errors Made by all Subjects Working on the Standard Program and on the Abridged Program from the End of Series 2SCD to the First Group of 10 Consecutive Correct Responses in Series 2SC

| *Errors* | *Frequency* | | | |
|---|---|---|---|---|
| | *Standard Program* | | *Abridged Program* | |
| | *Exploratory Exp. V* | *Control Group S* | *Exploratory Exp. III* | *Control Group A* |
| 0 | 1 | 2 | | |
| 1 | 2 | | 1 | 1 |
| 2 | 1 | 1 | 1 | |
| 3 | | | | |
| 4 | | | | |
| 5 | | 1 | 1 | |
| 6 | 1 | 1 | | 1 |
| 7 | 1 | | | |
| 8 | | | | |
| 9 | | | | |
| 10 | | 1 | | |
| 11 | | | 1 | |
| 12 | | | 1 | 1 |
| 13 | | | | 1 |
| 14 | | | | |
| 15 | | | | |
| 16 | | | | |
| 17 | | | | |
| 18 | | | | |
| 19 | | | | |
| 20 | | | | |
| 21 | | | | 1 |
| 22 | | | | 1* |
| 23 | | | 1 | |
| 24 | | | | |
| 25 | | | | |

* Subject did not reach criterion before the end of the series.

The contingencies of reinforcement arranged by the apparatus may have been somewhat unstable. Pressing a window required very little effort, and an incorrect response could be followed immediately by a correct one. It may have been easier to respond to position cues or to the properties of the choice stimuli and to switch after errors, accepting 50 per cent reinforcement, than to observe the matching relationship. In this case, the Standard Program may simply have prevented subjects from discovering an easier way to operate the apparatus. None of the subjects in Group S, for example, found out that the slide changer would not operate if both windows were pressed at once, because none of them tried it. However, the records of Group C in Session 2 did not show specific stable patterns of responding, as would be expected if subjects had hit on an easy way of operating the machine. Instead, several patterns alternated with one another. A more reasonable explanation of the performance of Group C on the Standard Program is that response patterns established under intermittent reinforcement in the Criterion Program conflicted with the observation of the relevant stimuli.

Randomizing the relationship of irrelevant stimuli to reinforcement guarantees that responses to them will be intermittently reinforced. In effect, a subject who is responding to a stimulus which has a 50 per cent correlation with reinforcement is working on a variable-ratio schedule with a mean ratio of two responses. In an insoluble discrimination, one might expect subjects to respond to one irrelevant stimulus and then to another, in the same way that animals alternate between responses maintained on equally profitable concurrent schedules (Herrnstein, 1958); but one would expect that these irrelevant stimuli would acquire substantial control which would conflict with the establishment of control by a new stimulus. The form of the records in Fig. 8 supports this interpretation. Only Subject 6 performed at a chance level from the beginning of the session. In the records of the other subjects who failed to learn the matching task, errors do not gradually increase, but occur as sharp discontinuities in the curves (marked by the letter "D" in the records). This is what would be expected if strong behavior intruded when the stimuli which controlled it appeared.

If deleting the transition series made it less likely for subjects to observe the relevant stimuli in the series that remained, one would expect subjects in Group A to make more errors, on the average, than those in Group S, and their performances would be more variable. The data from these two groups support this interpretation (Fig. 7 and 9). As would be expected, discontinuities appear in the records of subjects in Group A at the divisions between series, but they do not appear at the same divisions for all subjects; and the number of errors occurring before the curves level out again is quite variable. In contrast, the errors made by Group S were spread more evenly across the first session, except for the group of errors made by Subject 6 in Series 2SC. The points where errors occurred were not correlated with each other or with the steps in the program.

A clearer picture of the increase in variability caused by deleting the transition series is shown in Table 2, where the number of errors made in reaching a criterion of 10 consecutive correct responses on Series 2SC is compared for subjects working on two "abridged programs" and on two "standard programs."

The performance of Group A on the Criterion Program was surprising. Most of the errors made by Subjects 1, 2, 3, and 4 occurred after considerable periods of correct responding. The same phenomenon appears in the records of Group E (Fig. 10) in a more pronounced form, and to a somewhat lesser extent in Group R (Fig. 12). There is no immediately obvious reason why the number of errors occurring *after* learning should be a function of the training procedure.

In keeping with the interpretation of Group C's performance on the Standard Program, these "postlearning" errors may represent the intrusion of systematic forms of responding which had been incidentally strengthened during training. This explains why Group S made fewer postlearning errors than Group A: if the effect of the Standard Program was to make it more likely for subjects to observe the relevant stimuli and therefore less likely for "incorrect hypotheses" to be incidentally reinforced, then Group S had fewer "incorrect hypotheses" which were available to intrude later.

Two bits of evidence support this interpretation. (a) In general, the number of errors made by all subjects following the first substantial period of correct responding on the Criterion Program was correlated with the number of errors made up to that point. For subjects in the control experiment, the correlation between the number of errors made before the first group of 10 consecutive correct responses on the Criterion Program, and the number of errors made after that point was .63 (Pearson Product Moment Coefficient). (b) In the records of subjects who made a large number of postlearning errors, the errors both before and after learning tended to occur in groups, as would be expected if "incorrect hypotheses" were intruding.

This explanation is not entirely satisfactory. In the Criterion Program, the characteristic performance of nearly all subjects who learned the task was a more or less extended period of correct responding, followed by an increase in the slope of the error curve in which errors alternated with further periods of correct responding. In the records of subjects who made a large number of postlearning errors, this increase in slope often appears as a pronounced discontinuity in the curve. (Points where this discontinuity appears are marked "D" in Fig. 9 and 10.) A similar trend appears in many of the records of subjects in Groups S and S2 (Fig. 7 and 11), although these subjects made very few postlearning errors. If postlearning errors were caused by the intrusion of behavior which was incidentally reinforced during the learning period, one would expect the error curves in the entire criterion session to look like the portions which follow the discontinuities. What caused the curves to level out first?

The periods of errorless responding may have been due to an artifact in the procedure. They tended to appear at the beginning of the criterion session, and, in the records of Group S2, immediately after the experimenter reloaded the slide projector. The presence of the experimenter at these points may have acted as a supplementary source of motivation, which diminished as the session progressed. The record of Subject 4, in Fig. 10, shows the effectiveness of such supplementary motivation. At the beginning of the second session, the experimenter, contrary to the usual procedure, said, "This is an important job. I want you to work hard at it, with no nonsense." A considerable period of correct responding followed, which was broken by the clusters of errors appearing in the middle of the section. At "I," the experimenter said, "You're not paying attention; try to press the right one every time." Another period of correct responding followed, but the session ended with another cluster of errors.

Supplementary motivation due to the experimenter cannot account for periods of correct responding at the ends of the sessions, however. Another interpretation is attractive, because it helps to account for the inferior performance of Group E.

Three reasons why the operation of the slide projector acted as a reinforcer in these experiments are possible: (a) The subjects were told to "try to press the correct one every time," and they may have been working to please the experimenter or to reduce an implicit threat. (b) Operation of the slide projector resulted in stimulus change, and several studies have shown that such stimulus change may act as a reinforcer (*cf.* Berlyne, 1960). (c) Operation of the slide projector resulted in changes in *stimulus control.* Responding correctly under conditions of changing stimulus control possibly is reinforcing in itself. In a situation in which social motivation is minimized, as in these experiments, the effects of the latter two reinforcers should reveal themselves. Whether one or both of them were operating (and it may not be possible to separate the two), the effects would be the same. In a given series of slides, observing and responding to the relevant stimuli are reinforced by the presentation of novel items. As responding continues, the stimuli (and the task) become less and less novel, and observing should decrease. One would expect "careless" errors to occur, and eventually the subject should stop responding or respond to other properties of the apparatus. If supplementary motivation is added at this point, one would expect observing to occur as a kind of "avoidance response," maintained by occasional errors which result in punishment or loss of reward. On the other hand, if the series gives way to another, slightly different set of stimuli (and slightly different task), observing and correct responding should be maintained at a high level, reinforced by the appearance of novel items. The most effective moment to switch from one series to the next would be when correct responding is firmly established, but while observing is still at a high level, so as to maximize the chances that the subject will respond correctly to the new stimuli. If changes were made after observing had partially extinguished, they might go unnoticed for a time; and when observing did occur, the chances that the new stimuli would elicit an "incorrect hypothesis" would be increased. The subject might learn the task required by the new series; but once he was responding correctly, observing the relevant stimuli should again extinguish, and behavior which had previously been incidentally reinforced should reappear, making it less likely that the subject would notice the next change. The process would be cumulative, and the records of the subjects should show increasingly large groups of errors interspersed with periods of correct responding. This is a reasonably adequate *post hoc* description of the behavior of the subjects in Group E.

A similar explanation accounts for discontinuities in the records of subjects working on the Criterion Program. Subjects responded correctly until the monotony of the material caused responding to relevant stimuli to extinguish so that other behavior intruded.

The records of the subjects in Group R show that changes from one series to the next can take place too rapidly, as well as too slowly. In this case, the effect seems to be similar to that of the Abridged Program: Errors tend to occur in groups, associated with the points of transition between series; and a large number of errors occurs on the Criterion Program. Why longer periods of correct responding do not appear on the Criterion Program remains unexplained.

The preceding account is quite speculative, but it suggests several lines of research. How much the results were due to unstable contingencies of reinforcement can be assessed by varying the amount of effort involved in responding, for example, by requiring the subjects to respond to the correct window more than once in order to operate the apparatus. The effects of "extrinsic" motivation can be investigated by adding a score-counter or candy dispenser, or by causing errors to turn the apparatus off and requiring subjects to work to turn it on again.

The most interesting possibility is that responding correctly under changing stimulus control is automatically reinforcing. A particularly good way of studying this effect might be to switch individuals from one series to the next after a given number of responses at a given percentage of error and to study their behavior as a function of the criterion run.

## IMPLICATIONS

These experiments confirm the findings of Lawrence (1952), North (1959), and Spiker (1959) that learning is more efficient, in terms of either trials or errors, when training in a difficult discrimination is preceded by training in similar but easier discriminations, than when all training is given in the difficult discrimination alone.

Lawrence (1955) pointed out that these findings pose a problem for theories of discrimination learning: "Usually it is assumed that a fixed amount of practice is most profitably spent if all of it is devoted to training on the specific task on which the individual will be tested." He went on to show that the findings are difficult or impossible to account for in terms of generalization gradients.

However, Lawrence did not mention that his paradox rests upon the assumption that discrimination learning is "continuous." What to expect from a "fixed amount of training" on the "task on which the individual will be tested" is the question raised in the continuity-discontinuity controversy (*cf.* Blum & Blum, 1949).

McCulloch and Pratt (1934) gave the continuity-discontinuity problem its first, and probably its clearest, statement:

"The assumption seems to be implicit in nearly all reports upon sensory discrimination in the white rat that some quantitative aspect of the animal's performance preceding the solution, usually represented in terms of trials, errors, or time, reflects directly the capacity of the animal to make the discrimination.

". . . It is conceivable that the point at which an animal adopts the correct solution is dependent upon chance factors, *i.e.*, factors extrinsic to the animal's ability to form the discrimination concerned . . .

"If Lashley were correct in his interpretation that 'the practice preceding . . . is irrelevant to the actual formation of the association,' the measurement of a rat's ability in terms of the amount of practice preceding the solution would obviously be subject to criticism . . .

"If . . . it can be demonstrated that repetitive training produces a cumulative effect irrespective of the 'hypothesis' being tested, a summation of scores may be taken as representative of an animal's ability on a given task" (McCulloch & Pratt, 1934, p. 271).

Ehrenfreund (1948) showed that whether or not training produces a cumulative effect depends upon whether or not the experiment is designed so that subjects are likely to observe the relevant stimuli. The results of the present experiments may be considered an extension of Ehrenfreund's findings. If one accepts the argument developed above, that a program of successive discriminations makes it more likely for subjects to observe the relevant stimuli and therefore less likely for "incorrect hypotheses" to be incidentally reinforced, then the amount by which such a program increases learning efficiency must be a measure of the degree to which learning the criterion task alone is "dependent upon chance factors."

The method of approaching a discrimination through a series of successive approximations is offered here as a way of taking the discontinuities out of any discrimination experiment. The procedure may not resolve the controversy of continuity vs. discontinuity, but it seems to reduce it to another set of questions:

(1) If one wishes to know how likely it is for subjects to observe certain stimuli or respond to certain relationships—*e.g.*, do birds respond to visual stimuli more readily than rats?—then the classical discrimination experiments referred to by McCulloch and Pratt are valid.

(2) If one wishes to know whether subjects are *capable* of responding to a given relationship—*e.g.*, the "conditional discriminations" described by Lashley (1938) or the present matching-to-sample task—it would seem more reasonable to approach the criterion task by a series of successive approximations and to observe where the performance breaks down. In the present experiments, for example, one might expect a subject "incapable" of responding to the matching relationship to begin making errors as the sample was moved to a point midway between the choice stimuli.

(3) If one is interested in how subjects go about discovering which of a given set of stimuli is correlated with reinforcement or in the capacities of subjects to solve such a problem, an experiment in which subjects are first trained to observe the component stimuli and then are required to "discover the concept" is indicated (*cf.* Harlow, 1949).

(4) One may be interested in whether or not discrimination learning is "continuous," in the sense that more than one reinforcement is necessary to "confirm a hypothesis," and to what extent some individuals learn more quickly than others, *i.e.*, require fewer reinforcements. Some of the data from the present experiments relate to the first of these questions. The effect of the Standard Program lends itself to a Guthrian interpretation (Guthrie, 1935); *i.e.*, the program was effective because gradual changes in stimulus control insured that new stimuli were introduced in such a way as to prevent incorrect responses from being elicited. If this were a matter of strict contiguity, shortening the series should have had little effect, so long as each of the items in each series appeared at least once. The performance of Group R suggests that this is not the case, but the question should be pursued further.

These experiments raise a related question. Discrimination is ordinarily described as a process of differential reinforcement (*cf.* Spence, 1936; Skinner, 1938; Hull, 1950). If responses to the negative stimulus are strong, whether through generalization from the positive stimulus or through prior training, the only mechanism which can weaken them, according to this interpretation, is nonreinforcement (disregarding punishment). The fact that errors occurred when subjects working on the Abridged Program were switched from Series 2SC0 to Series 2SC would seem to indicate that the negative stimuli in Series 2SC had acquired some control over the choice-response through training in Series 2SC0. There is no clear reason why gradually increasing the intensity of these stimuli, as was done in the Standard Program, should have weakened their control. When they reached their full intensity, they should have elicited a proportionate number of errors. The fact that they did not do so, and that 5 of the 18 subjects who worked on the Standard Program made no errors on Series 2SC, is rather surprising. Subject 1 in Group 6 and Subject 1 in Exploratory Experiment 5 learned the task required by the Criterion Program without making an error. In subsequent experiments, this phenomenon has become commonplace. These children learn the matching task without discovering that pressing an incorrect window does not operate the machine.

One purpose of these experiments was to see how practical an apparatus of this sort might be for teaching discriminative and conceptual behavior to children. Two related findings appear to have special practical importance.

(1) The more errors the training procedure allowed children to make, the more they tended to go on making. What would happen if the subjects who learned the task required by the Criterion Program were all given a further program—for example, one in which the number and similarity of the stimuli were gradually increased? This has not yet been tested, but it seems likely that differences in performance will be related to differences in performance on the Criterion Program. In the present situation, there appears to be a progressive loss of control. After a bad program has been tested, it can be revised, but the subjects must be thrown away. Additional "extrinsic" motivation may minimize post-learning errors for all children, regardless of the training procedure, but this remains to be investigated.

(2) Programs can be too long as well as too short or too discontinuous. It may be possible to come empirically to programs which are an effective length for most children; the Standard Program was reasonably adequate in this respect. Again, additional sources of motivation may make the performances of

subjects more consistent, both by reducing the number of postlearning errors and by decreasing the aversive effects of the program for those subjects for whom it is too long. On the grounds of efficiency, however, one would prefer to let subjects proceed to the next task as rapidly as possible after mastering the first. Requiring a child to work through truly repetitive material beyond the time when he could effectively pass on to the next task is educational "lock-step" in a teaching machine.

Operant-conditioning studies with children (*cf.* Bijou, 1955, 1957; Long, *et al.*, 1958) have shown how hard it is to maintain behavior on a repetitive task. The apparatus for programming additional reinforcement or punishment in the present discrimination task may be both elaborate and unnecessary. An abundance of anecdotal evidence suggests that the reinforcing effect of responding correctly under conditions of changing stimulus control may be real. Teachers who have worked with individual students, and parents who have played with small children—setting them problems which differ just enough to "challenge" them—will be the first to affirm the motivating effect of the procedure. If such an effet is found in the present situation—and the way to find it is by switching subjects to the next series after some some given criterion run—it will be a truly "intrinsic" motivation. It will not be surprising if it turns out to be a powerful one.

## REFERENCES

Berlyne, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

Bijou, S. W. A systematic approach to an experimental analysis of young children. *Child Develpm.*, 1955, **26,** 161-168.

Bijou, S. W. Methodology for an experimental analysis of child behavior. *Psychol. Rep.*, 1957, **3,** 243-250.

Blum, R. A., and Blum, J. S. Factual issues in the continuity controversy. *Psychol. Rev.*, 1949, **56,** 33-50.

Ehrenfreund, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.*, 1948, **41,** 408-422.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.* 1933, **42,** 207-208.

Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.

Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56,** 51-65.

Herrnstein, R. J. Some factors influencing behavior in a two-response situation. *Trans. N.Y. Acad. Sci.*, 1958, **21,** 35-45.

Holland, J. G. Teaching machines: an application of principles from the laboratory. *J. exp. Anal. Behav.*, 1960, **3,** 275-287.

Hull, C. L. Simple qualitative discrimination learning. *Psychol. Rev.*, 1950, **57,** 303-313.

Krechevsky, I. 'Hypotheses' in rats. *Psychol. Rev.*, 1932, **39,** 516-532.

Lashley, K. S. Conditional reactions in the rat. *J. Psychol.*, 1938, **6,** 311-324.

Lawrence, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45,** 511-516.

Lawrence, D. H. The applicability of generalization gradients to the transfer of a discrimination. *J. gen. Psychol.*, 1955, **52,** 37-48.

Long, E. R., Hammack, J. T., May, F., and Campbell, B. J. Intermittent reinforcement of operant behavior in children, *J. exp. Anal. Behav.*, 1958, **1,** 315-340.

McCulloch, T. L., and Pratt, J. G. A study of pre-solution in weight discrimination by white rats. *J. comp. Psychol.*, 1934, **18,** 271-290.

Mostellar, F., and Bush, R. R. Selected quantitative techniques. In Lindzey (Ed.), *Handbook of social psychology*, Vol. 1. Cambridge: Addison-Wesley, 1954. p. 315.

North, A. J. Acquired distinctiveness of form stimuli. *J. comp. physiol. Psychol.*, 1959, **52,** 339-341.

Skinner, B. F. *The behavior of organisms.* New York: D. Appleton Century Co., 1938.

Skinner, B. F. 'Superstition' in the pigeon. *J. exp. Psychol.*, 1948, **38,** 168-172.

Skinner, B. F. The science of learning and the art of teaching. *Harvard educ. Rev.*, 1954, **24,** 86-97.

Skinner, B. F., and Morse, W. H. A second type of superstition in the pigeon. *Amer. J. Psychol.*, 1957, **70,** 308-311.

Skinner, B. F. Teaching machines. *Science*, 1958, **128,** 969-977.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43,** 427-449.

Spiker, C. C. Performance on a difficult discrimination following pre-training with distinctive stimuli. *Child Develpm.*, 1959, **30,** 513-521.

White, B. W. Complexity and heterogeneity in the visual recognition of two-dimensional forms. In J. W. Wulfeck and J. H. Taylor (Eds.), *Form discrimination as related to military problems.* Washington, D.C.: Nat. Acad. Sci.-Nat. Res. Council, 1957. p. 158. 158.

*Received June 28, 1961*

# PARAMETRIC EXPERIMENTS WITH A MATCHING-TO-SAMPLE PROGRAM

## Wells Hively

In some earlier experiments (Hively, 1962) kindergarten and first grade children were put through a program of progressively difficult discrimination tasks, leading up to matching-to-sample(see previous chapter). Such training was found to be much more efficient than training in the final discrimination alone.

The apparatus used in the above experiments is discussed earlier (Hively, 1964) [see page 67 ]. Items from the program are shown schematically in Fig. 1. A random sequence of the four items in rows 9 and 10 of Fig. 1 constituted the final matching-to-sample task. The preceding rows represent steps in the program which led up to it. In Steps 1 through 6 the positions of the choice-stimuli were held constant, the sample was located directly above the correct choice, and the incorrect choice was gradually "faded in." In Steps 7, 8, and 9, the sample was gradually moved to a point midway between the two choices. Finally, in Step 10 (represented by rows 9 and 10 in Fig. 1), the positions of the choice-stimuli were allowed to vary at random.

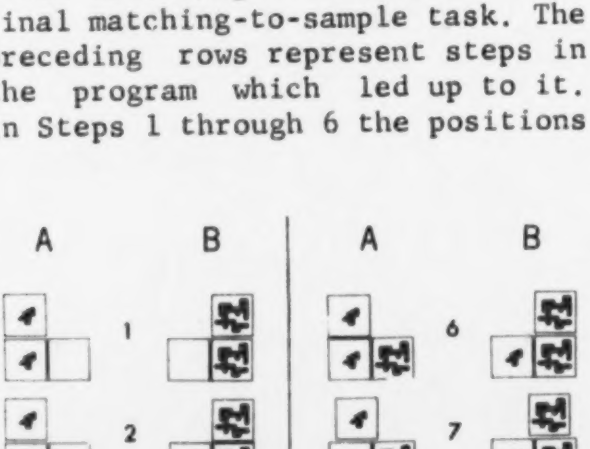


Fig. 1. Items in the matching-to-sample program.

Each step in the program may be thought of as a separate discrimination task--a random sequence of the items in any given step constituting a test of that particular discrimination. The main question raised as a result of the initial experiments was how many slides to put into each step. A "Standard Program"seemed to contain an optimum number of slides. Both "expanding" it,by doubling the number of slides in each step and "reducing" it, by halving the number of slides caused children to make more errors.

These findings suggested that switching subjects to the next step as soon as they made a given number of successive correct responses might be more efficient than putting them through a program of fixed length. In an exploratory experiment with such a contingency (Experiment 2,below) several children made no errors at a criterion of two--the equivalent of a fixed-program with only two slides in each step, and considerably shorter than the "Reduced Program"which seemed to cause more errors in the initial experiments. Consequently, the following studies were run as a systematic replication of the original ones, with the number of slides in each step as a parameter.

## EXPERIMENT I

### Varying the Number of Slides in Each Step of the Program.

Apparatus. The apparatus and procedure have previously been described in detail (Hively, 1962, 1964).

Programs. As in the earlier experiments, a random sequence of the four items in rows 9 and 10 of Fig. 1 constituted the "Criterion Task." Calling the items "9A" (left column), "9B" (right column), "10A" (left column), and "10B" (right column)--the test consisted of the following twenty slides: 9A, 9B, 9A, 10B, 10A, 10A, 9B, 9A, 10B, 9B, 9B, 9A, 9A, 10B, 10A, 9B, 10B, 10A, 10A, 10B. This series was presented in the order shown. After the 20th slide the projector reversed automatically and presented the series in backward order, turning around again at Slide 1. This cycle could be repeated indefinitely.

To compose the programs, a Gellerman Series (Gellerman, 1933) was used to determine the column from which each successive item was selected (Fig. 1), and a given number of items was taken from each row (i.e., step in the program) up to and including Row 9. The first two slides in Step 10 were always the items 10A and 10B, and thereafter the sequence of slides was the same as in the Criterion Task. Seven versions of the "Standard Program" were composed in this fashion, together with six versions of the "Abridged Program", which included only Steps 1, 6, 9, and 10. Table 1 shows the number of slides in each step of these programs (not counting the four slides used by E to demonstrate the task).

Subjects. Subjects were kindergarten and first grade children in suburban schools. They were assigned to experimental groups in such a way as to distribute possible effects of age, sex, and classroom. Their teachers were asked to discuss the experiment with them as little as possible. Six first grade and six kindergarten children were run on each program. (The experiments were carried out in three different schools at intervals throughout the second

Table 1

Programs with an Equal Number of Slides in Each Step

| Standard (10 step program) | | Abridged (steps 1,6,9, and 10 only) | |
|---|---|---|---|
| Number of slides in each step | Total in program | Number of slides in each step | Total in program |
| 1 | 11[a] | 2 | 8 |
| 2 | 20 | 4 | 16 |
| 4 | 40 | 6 | 24 |
| 6 | 60 | 10 | 40 |
| 9 | 90 | 15 | 60 |
| 12 | 120 | 30 | 120 |
| 22 | 220 | | |

[a] Step 10 contained both items shown in Row 10 of Fig. 2.

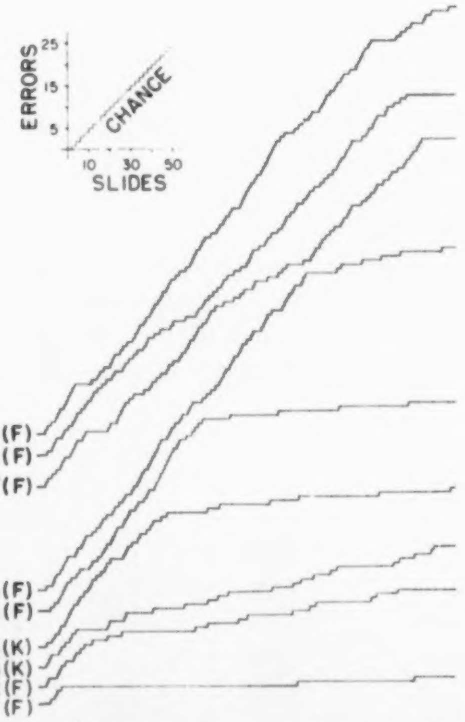


Fig. 2 Records of the 9 best subjects out of 22 who worked only on the criterion task.

half of the school year.) After completing a program, each child immediately worked through 120 slides of the Criterion Task. In addition, 22 children worked through 200 items of the Criterion Task without having first received a program.

Results. Records of the nine best subjects out of the 22 who worked only on the Criterion Task are shown in Fig. 2. The records were drawn by plotting errors cumulatively against the series of 200 slides. Changes in the slopes of the curves show changes in the accuracy with which the subject responded: with the scales used, chance performance yields a slope of approximately 45 degrees (see inset). During periods when a subject was making no errors, his curve is a horizontal line. The symbols (F) and (K) identify the subjects as first grade or kindergarten children.

Note that transitions from incorrect to correct performance appear as relatively sharp discontinuities in the curves, and the slopes before and after the points of inflection are clearly different. The numerical criterion which best locates these points of inflection appears to be a dual one: either 10 consecutive correct responses or a sequence of 20 slides including no more than 3 errors and beginning with a run of 6 consecutive correct responses. Throughout the following discussion the term "reaching criterion" refers to this measure.

Records of the children working on the Standard Programs and on the Abridged Programs are shown in Figs. 3 and 4 respectively. The groups are identified by numbers at the upper left which refer to the number of slides in each step in the program (to conserve space, the groups are not arranged in order). Performance on the programs and the subsequent Criterion Task are plotted continuously but the two are separated by a small gap. Steps in the programs are marked by the "staircase" symbols above and below each group: errors may be roughly located by setting a straight edge to these symbols and observing where the curves cross it. In order to conserve space, several records have been cropped off at the top--all of them continued at a chance slope throughout the Criterion Task. (Note that some children in the Abridged Groups received only 25 slides of the Criterion Task.)

The most striking characteristic of these data is the variability within groups. Every group includes one or more subjects who made few or no errors (this is true of the groups working on the Abridged Programs as well as those working on the Standard Programs), and every group, with the exception

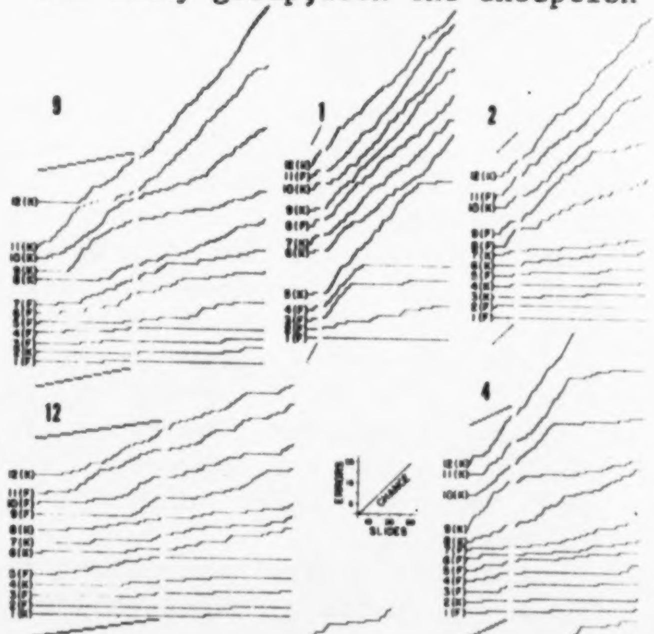


Fig. 3. Records of subjects who worked on the standard program. The numbers above the groups refer to the number of slides in each step program.

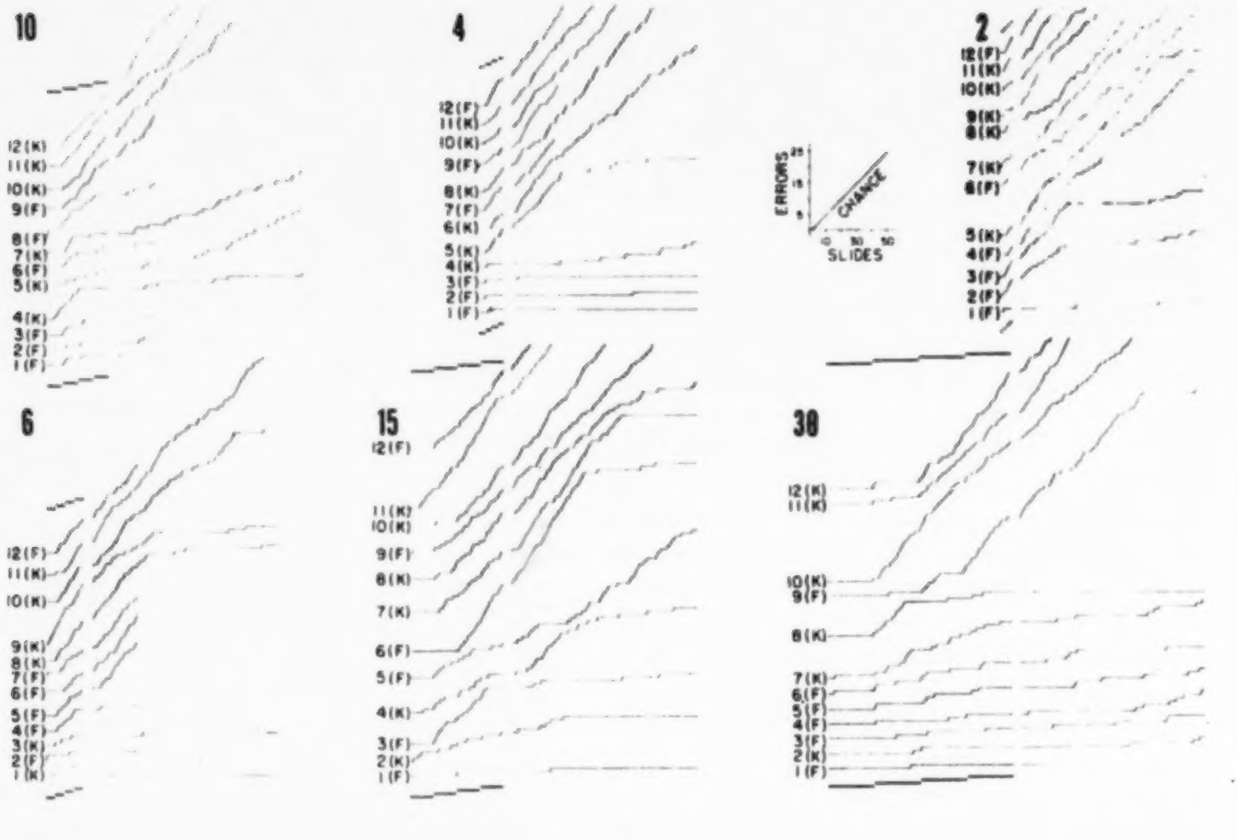


Fig. 4. Records of subjects who worked on the abridged program. The numbers above the groups refer to the number of slides in each step of the program.

of Standard Groups 12 and 22, also includes one or more subjects who failed to learn the task at all. The children tended either to be under control of the discriminative stimuli throughout the session or else they tended to be out of control over a large number of slides. The main effect of the programs, whatever their length, was to increase the number of subjects who were under control from the beginning.

There are several exceptions to the above pattern. Some children continued to make errors after they reached criterion, and these "post-learning" errors tended to come and go in bursts, giving their curves a pronounced "grain" and an intermediate slope. In the shorter programs, records of this sort are exceptions, but in Standard Groups 9 and 12 they are the rule. In general, children who reached criterion after extended runs at a high error-rate tended to make few errors thereafter, and children who reached criterion with very few or no errors continued to make few. However, children who reached criterion after short runs of errors at intermediate overall rates tended to go on making errors.

In Figs. 5 and 6 the percentage of subjects who reached criterion in each group is cumulated against the number of slides to which they responded. The group of children who worked only on the Criterion Task is shown in both figures by open circles. Three of the 22 subjects reached criterion within the first 30 slides, another reached criterion on Slide 65, and so on. Nine subjects, or 40% of the group, had reached criterion by the time they had responded to 190 slides. The curve stops at 190 because no further subjects reached criterion within the 200 slides.

Records of the groups working in the Standard Programs are shown in Fig. 5, and those of the groups

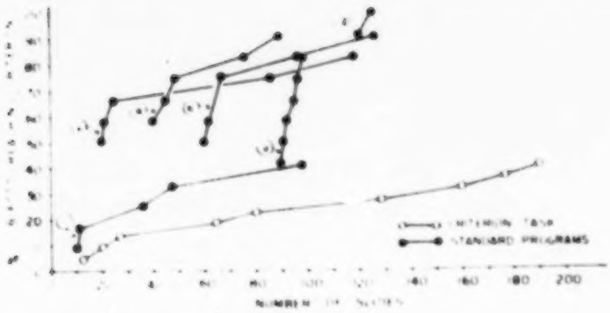


Fig. 5. Group curves for standard programs of various lengths. The percentage of subjects who reached criterion in each group is cumulated against the number of slides to which they responded.

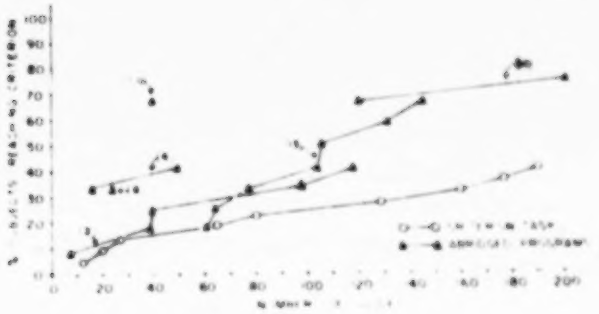


Fig. 6. Group curves for abridged programs of various lengths. The percentage of subjects who reached criterion in each group is cumulated against the number of slides to which they responded.

working on the Abridged Programs in Fig. 6. The first point on each curve represents the percentage of subjects who reached criterion immediately upon being presented with the Criterion Task. (Although they may have made no errors on the program, none could be said to have reached criterion sooner.) Thereafter, the curves are plotted in the same way as that of the Criterion Group. All curves but that of the Criterion Group are based upon 12 subjects, so the points may be read in terms of absolute numbers, as well as percentages, of subjects.

The relative effectiveness and efficiency of the various programs may be judged by the location of the curves: curves lying in the lower portion of the figure represent programs which were relatively ineffective; curves lying in the upper right represent programs which were effective, but relatively inefficient; and curves lying in the upper left represent programs which were both efficient and effective.

The curve of the Criterion Group has a steady positive slope, showing that training in the Criterion Task alone, although relatively inefficient, had some effect. The curves of Standard Groups are all sharply inflected. Note that all of them, following the points of inflection, are more or less parallel to each other and to the curve of the Criterion Group. In other words, children who did not reach criterion very soon after completing a program behaved very much as if they had never had one. The points of inflection of the curves also fall along a line which is more or less parallel to the curve of the Criterion Group. That is, the effect of increasing the number of slides in each step of the Standard Program, beyond a minimum of two in each step, was approximately the same as that of an equal amount of practice devoted to the Criterion Task alone. However, reducing the number of slides in each step from two to one substantially destroyed the program's effect.

One might expect an interaction between the "size of the steps" in in each step. While large stimulus changes cause subjects to make errors, a large number of slides in each step might make it possible for them to recover before the next stimulus change. However, increasing the number of slides in each step appears to have had no more effect upon the Abridged Groups than upon the Standard. The variability of the performances on the Abridged Programs makes it hard to detect a trend.

Standard Program 12 contained the same total number of slides as the original "Standard Program" (Hively, 1962). Similarly, Standard Program 9 may be compared to the original

"Reduced Program"and Standard Program 22 may be compared to the original "Expanded Program". The performances of subjects on the first two pairs of programs were fairly similar, although there was less difference between Standard Program 12 and Standard Program 9 than between the original Standard and Reduced Programs. However, the performances of subjects on Standard Program 22 was much better than those on the original Expanded Program. These two programs differed in the proportion of slides given to each step. To check wether this was what caused the errors, another version of the Expanded Program--"Expanded Program 11"--was composed: it was the same in over-all length, but steps 2,3, 4, 5, 7, and 8 each contained only four slides, while steps 1, 6, 9, and 10 each contained 55. The performances of subjects on this program are shown in Fig. 7. Altering the proportions of slides assigned to each step did have a disruptive effect. Note the intermediate slopes in the records of subjects 7, 8, and 10.

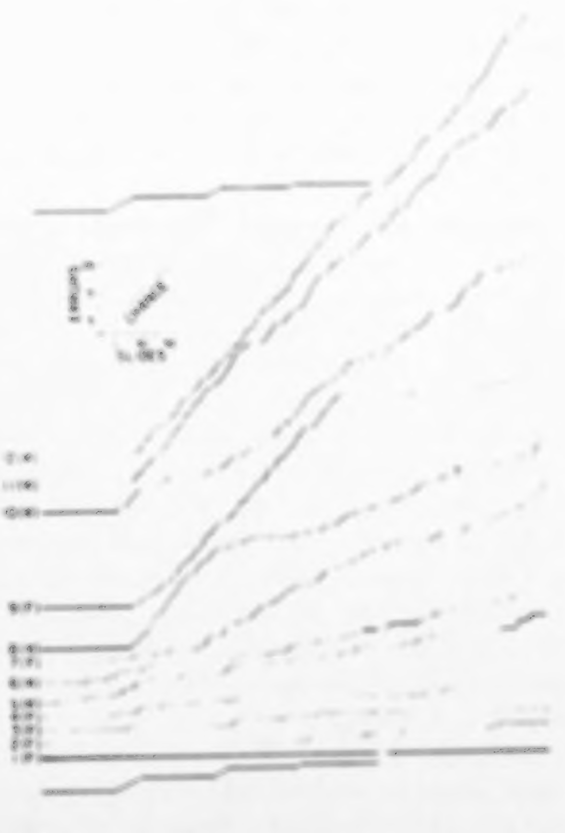


Fig. 7. Records of subjects who worked on Expanded Program 11.

Discussion. The Standard Program was most efficient when it contained only two slides in each step. Decreasing the number from two to one destroyed its effectiveness, while increasing the number had a relatively small effect. Nevertheless, the longer the program, the better was the performance. "Boredom" had no detectable effect on the accuracy with which the children responded. This finding is contrary to my prediction on the basis of the earlier experiments.

Performances on the Abridged Programs were similar to those on the Standard Programs,except that fewer children were brought under control,and there was more variability between groups. There was no indication that additional practice could compensate for larger steps.

Individual records tended to fall into these patterns:

1. Few or no errors throughout.
2. Long runs of responding at a chance level,followed, after sharp discontinuities, by few or no errors.
3. Short bursts of errors, interspersed with short stretches of correct responding, giving the records an overall intermediate slope.

In general,the longer the program, the more records fell into the third pattern. There is a plausible explanation for this: the conditions which seem to produce intermediate slopes are those in which the discriminative stimuli change slowly enough that if control is lost the subject does not go rapidly into debt. In Standard Programs 1, 2, and 4, the stimuli changed so rapidly that when control was lost it tended to be lost completely. The stimuli in Standard Programs 9 and 12 changed slowly enough that stimulus control might be repeatedly lost and re-established: irrelevant stimuli could perhaps acquire some strength in this process and continue to elicit "post-learning" errors.

It would follow from the above that introducing small discontinuities into a long program should produce some records with intermediate slopes. This appears to have been the effect of Expanded Program II (Fig. 7) as compared to Standard Program 22 (Fig. 3).

Records falling into the second pattern, however, do not fit this explanation. In general, if subjects learned the task at all after a long run of chance responding, their records showed an abrupt transition to sustained accurate performance. Since there was ample opportunity for irrelevant stimuli to acquire some control, why didn't these children make more "post-learning" errors?

The results of these experiments build a strong case for the "discontinuity" of the learning process in this particular discrimination task. The discontinuity is shown in the sharp transitions from chance to errorless performance on the Criterion Task and in the equally sharp transitions from errorless to chance performance during the course of a program. It is also reflected in the fact that the program required only a small number of slides in each step in order to achieve its main effect. It is difficult to think of a "step" as a separate "discrimination task" when it includes only two slides. It seems more reasonable to think of the program as a sequence of items designed to transfer very specific stimulus control, rather than as a series of "successive approximations" to the final discrimination task.

## EXPERIMENT II

### Run-to-Criterion in Each Step of the Program.

Method. A "Standard Program" was composed which had ten slides in each step. The slide projector cycled back and forth within each group of ten slides, presenting the series forward and backward indefinitely, until a given number of consecutive correct responses occurred: when criterion was reached, the projector automatically moved to the next step and began to cycle through the next series of ten slides. (In moving to a new step the projector momentarily showed all the intervening slides and occasionally this display lasted for several seconds.) The apparatus and procedure were otherwise identical to those in Experiment 1. Groups of six children were run at criteria of 1, 2, 3, 4, 5, and 9 consecutive correct responses. All of the subjects were first grade children.

Results. Records of subjects working under the run-to-criterion condition are shown in Fig. 8. The groups are identified by numbers which indicate the length of the criterion run. These records were plotted in the same way as those in the preceding experiments, except that in order to conserve space, the records of subject 6 in Group 3 and subject 6 in Group 5 have been set back to their base lines. Short vertical lines beneath the records mark the points of transition to each new step. Note that the Criterion Task was much shorter than that used in Experiment 1.

Requirements of 1, 2, and 3 consecutive correct responses were not sufficient to hold subjects in the same step when they began making

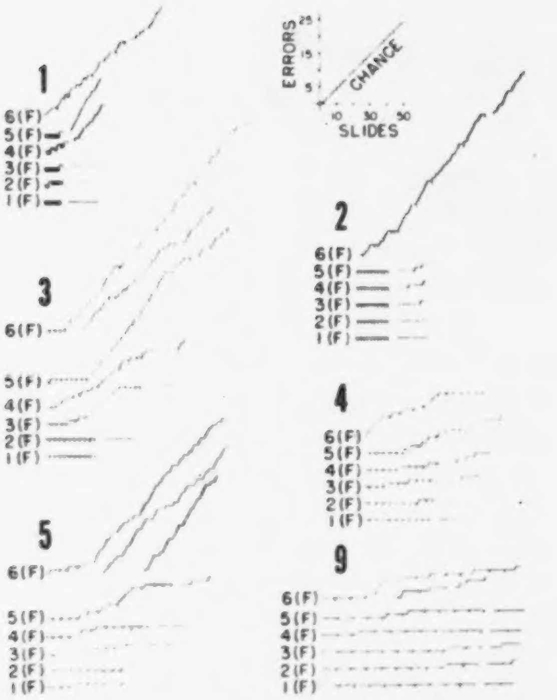


Fig. 8. Records of subjects who worked under the run-to-criterion condition. The numbers above the groups refer to the length of the criterion run.

errors. Note, however, that a criterion of 5 prevented Subject 6 from getting out of Step. 4.

Discussion. The performances of these subjects were more or less similar to those of subjects who worked on comparable fixed programs. Holding subjects in a given step when they began to make errors was not particularly effective. But, more important, a criterion large enough to hold subjects once they begin to make errors (about 4 consecutive correct responses) dictates a program twice as long as the optimum fixed program (two slides in each step). One might get around this difficulty by using dual criteria. For example, a subject might be presented with a program having two slides in each step, but if he made an error he might then be required to meet a criterion of 6 consecutive correct responses in order to leave that particular step. If he made no further errors, the next steps would again contain only two slides. Programs of this type have not been investigated.

## EXPERIMENT III

### "Drive-Back"

What accounts for the discontinuities in the records of subjects in Experiments 1 and 2? Can the process be made more continuous? From a practical point of view, what do you do with a child when he begins to make errors? The results of Experiments 1 and 2 showed that holding him in the same step until he comes back under control is not particularly effective. What about returning to an earlier step?

Subjects may be presented with a relatively short program, containing a fixed number of slides in each step, with the provision that they go back to an earlier point each time they make an error. They may be required to return to a particular slide (i.e., the beginning of the preceding step), or they may simply be required to go back to a given number of slides from the point at which the error was made. Under the latter condition there is an interesting relationship between the distance which the subject is set back and the probability of his emitting a correct response by chance. In a two-choice discrimination task, for example, the probability of chance success is .5, and therefore a contingency requiring subjects to go back one slide for each error should keep the subject who is out of control in roughly the same place. In a program which required subjects to go back two slides for each error, he would gradually drive himself back to the beginning of the program. Larger "drive-back" values would push him back faster.

Method. As always, each time a correct response was made the shutter closed and the slide projector operated. Whenever a correct response was not preceded by an error, the projector presented the next slide in the program. However, whenever a correct response was preceded by an error, the slide projector reversed and counted backwards for a certain number of slides. ( The shutter remained closed until the projector came to a stop on the proper slide. ) In "drive-back 1" the next slide presented to the subject was the one immediately preceding the slide upon which he made an error and in "drive-back 2" the slide presented was two slides before the one upon which he made an error.

The overall effect of the drive-back contingency may be judged from Fig. 9, where group curves similar to those in Fig.6 are plotted, but in terms of absolute numbers, rather than percentages, of subjects. The curve for the group of subjects who worked on the Standard Program with two slides in each step in Experiment 1 is included for comparison. Drive-back 1 had no effect. Drive-back 2 smoothed the group curve, but the over - all performance was the same.

Groups of subjects were also run on the Standard Program with one slide in each step at drive-back 1 and drive-back 2. Group records are shown in Fig.10, together with the records of groups working on the Standard Programs with one and two slides in each step in Experiment 1.

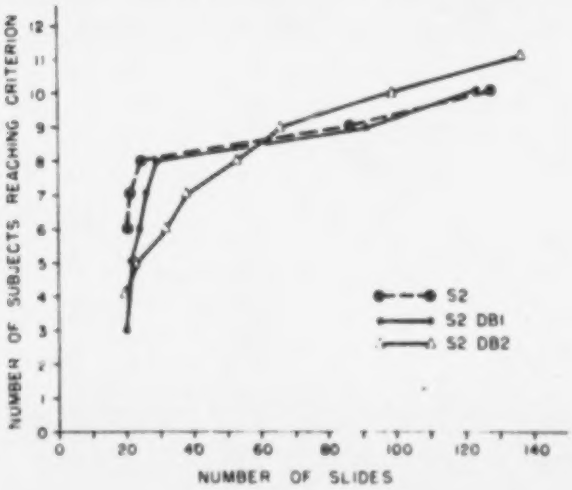


Fig. 9. Group curves for drive-back 1 (DB1) and drive-back 2 (DB2) on the baseline of the standard program with two slides in each step (S2). The number of subjects who reached criterion in each group is cumulated against the number of slides to which they responded.

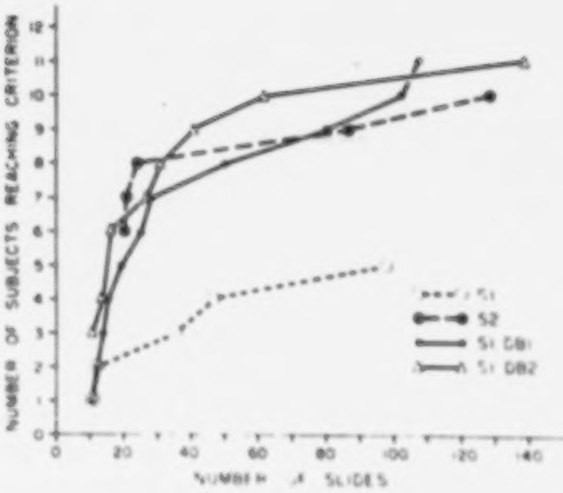


Fig. 10. Group curves for drive-back 1 (DB1) and drive-back 2 (DB2) on the baseline of the standard program with one slide in each step (S1). The number of subjects who reached criterion in each group is cumulated against the number of slides to which they responded.

The drive - back contingency did have an effect on the performances of subjects working on the program which had only one slide in each step. Roughly speaking, the effect was to make it equivalent to the program which had two slides in each step. There was little difference between drive - back 1 and drive-back 2.

Discussion. Post-learning errors were quite frequent in the records of the subjects who worked under these conditions. They were somewhat more frequent in the records of subjects who worked on the program which had one slide in each step than in the records of those who worked on the program which had two. This might be expected. As Experiment 1 showed, the Standard Program with one slide in each step was too short to set up adequate stimulus control. The drive-back contingency compensated for this deficiency and, therefore, stimulus control was being repeatedly lost and re-established. These are the conditions which might be expected to produce post-learning errors.

One could argue that returning

to an earlier point in the program may simply expose the subject again to the same sequence of stimuli which elicited the error the first time around. Some of the records show patterns which indicate that this sort of vicious cycle might have been set up. A more effective drive-back program might be one in which the over-all sequence of items is determined by the Gellerman Series, with the provision that after an error the subject is presented with the next item as determined by the series but taken from the same step as the slide just before the one on which he made the error (drive-back 1) or two slides before it (drive-back 2). This type of program has not yet been studied.

## EXPERIMENT IV

### Ratio Contingency

Several aspects of the children's performances in the foregoing experiments suggested that the contingency of reinforcement was not very precise. After an incorrect response, the subjects could respond again immediately, and under these conditions it was almost as easy to respond without looking and to switch after errors as it was to observe the matching relationship. Therefore, in the present experiment the following changes were made.

Method. A short fixed-ratio of responses to the window upon which the correct stimulus appeared operated the shutter and the slide projector. Responses to the incorrect stimulus reset the ratio but had no other effect (subjects could hear the ratio counter reset). Subjects received the usual instructions, during which they watched the experimenter "run out" the ratio on four consecutive slides.

The ratio contingency was intended to make it less likely that switching would be reinforced. In addition, the set-up had some other interesting features. Since no event was contingent upon completing the ratio when the choice was incorrect, the length of the run might serve as a measure of the strength of the incorrect response. Similarly, switching from an incorrect to a correct choice before the ratio had been completed might indicate partial stimulus control. The ratio might also filter out careless or playful responses and prevent them from being recorded as errors.

Results. The records of Group 1 in Fig. 11 show the effect of adding the fixed-ratio contingency (FR 3) to the standard program with one slide in each step and drive-back 2. (Compare Fig. 12, group 2). The effect of the ratio contingency appears to depend upon the extent to which the ratio performance had stabilized before the discriminative stimuli began to change. For Group 1E (Fig. 11) 10 slides were added to step 1. Expanding step 1 improved performances dramatically.

In these records, the programs were plotted somewhat differently in order to show the operation of the drive-back contingency. Take, for example, the record of Subject 4 in Group 1. This subject went

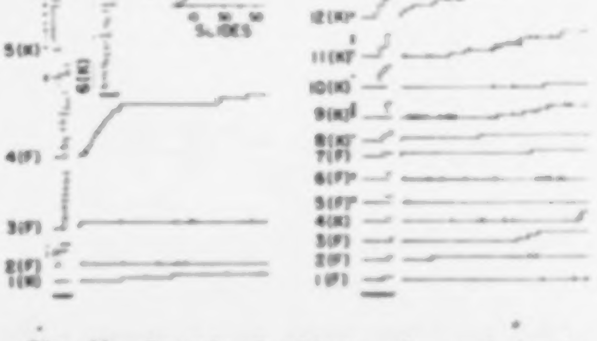


Fig. 11. Records of subjects who worked under the fixed-ratio condition and drive-back 2 on the standard program with one slide in each step (Group 1) and on the same program with an expanded first step (Group 1E).

over half-way through the program without making an error, as shown by the horizontal line beginning after the symbol 9(K). The line stops at the point where the subject made his first error. A new line begins at the slide to which the subject was driven back and a distance above the first line equal to the distance which the pen steps up when an error is recorded in the ordinary cumulative error record. Subject 9 made an error on that slide, too. Therefore a new trace begins above and behind the second one, and so on.

Notice that the records of subjects working under the fixed-ratio contingency are plotted in a special way. In the records for the Criterion Task in Fig. 11, incorrect responses which failed to complete the ratio are marked by vertical lines, but are not cumulated. Incorrect responses which met or exceeded the ratio are cumulated in the normal way. Switches from the correct to the incorrect stimulus are marked by dots beneath the curves: these errors are always cumulated, although the subject may have switched back to the correct stimulus after a short run. In the programs, "short-run" or "partial" errors are marked by arrows and switches from the correct to the incorrect stimulus are marked by dots.

If a subject were to respond with his eyes closed, one would expect that if he switched after short runs, the switches should go equally often in both directions. Notice the subjects in Group 1E switched from incorrect to correct choices after short runs much more often than they switched from correct to incorrect. This must represent either some degree of stimulus control or the effect of the ratio in filtering out careless responses. Therefore short-run errors were not cumulated in any of the records, but bear in mind that wherever a substantial number of short-run correct responses is found short-run errors should be thought of as cumulated.

Figure 12 shows the effect of the ratio contingency on perfor-

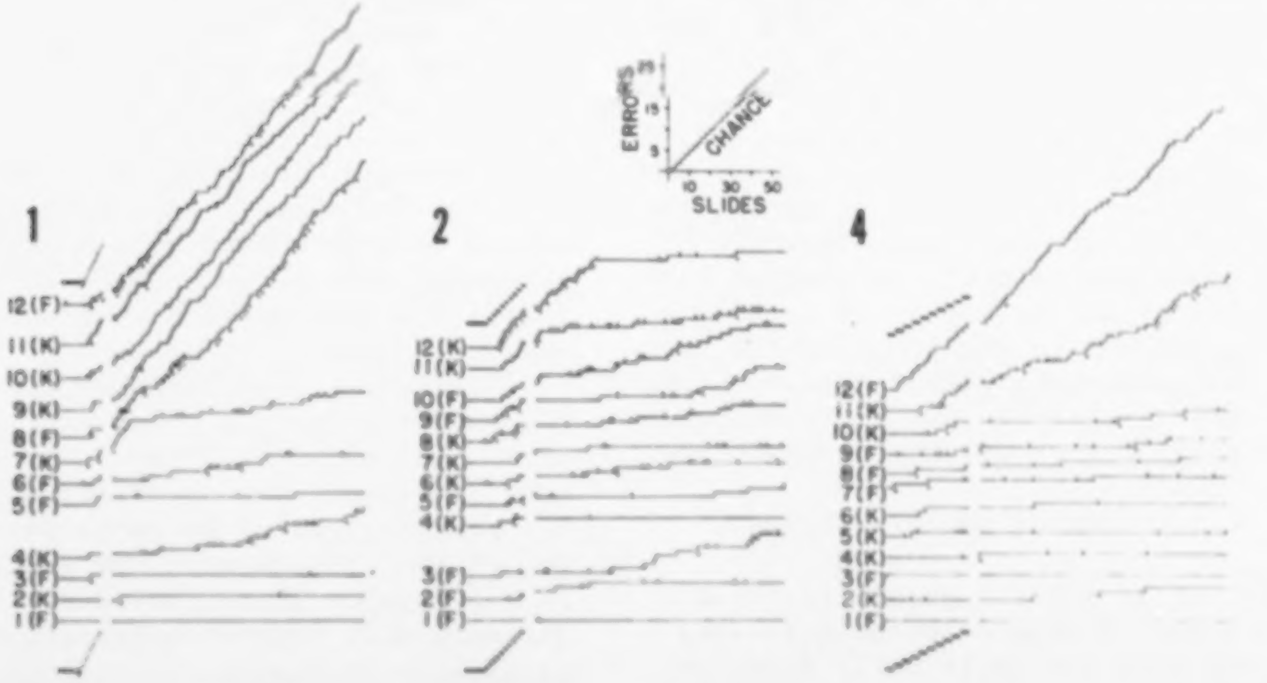


Fig. 12. Records of subjects working on the standard program with one, two, and four slides in each step under the fixed-ratio condition. The numbers above the groups refer to the number of slides in each step. (Programs 1 and 2 each had ten slides in the first step.)

mance in the Standard Programs with 1, 2, and 4 slides in each step. (Programs 1 and 2 were altered to include ten slides in Step 1.) The performances of all groups were considerably better than those of comparable groups in Experiment 1 (Fig. 3) but a clear difference remained between Standard Programs 1 and 2. Note that the ratio contingency did not, by any means, eliminate (or filter out) post-learning errors.

Figure 13 shows the group curves for Standard Programs 1 and 2, together with the curve for Standard Program 1, drive-back 2,--all with the ratio contingency and expanded first step. Again, the effect of the drive-back contingency was to make the program with one slide in each step roughly equivalent to the program with two. Note, however, that the fixed Standard Program 2 was still the most efficient of all the programs used.

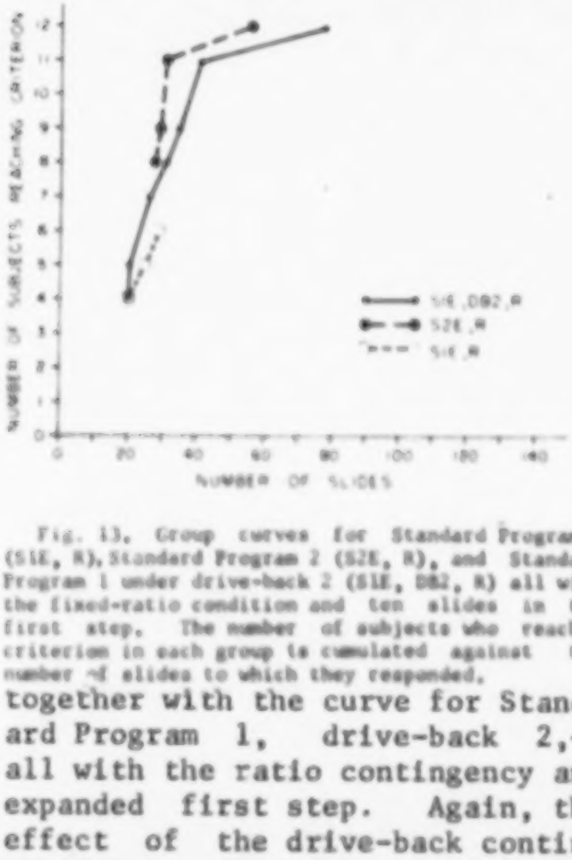


Fig. 13. Group curves for Standard Program 1 (S1E, R), Standard Program 2 (S2E, R), and Standard Program 1 under drive-back 2 (S1E, DB2, R) all with the fixed-ratio condition and ten slides in the first step. The number of subjects who reached criterion in each group is cumulated against the number of slides to which they responded.

## SUMMARY AND DISCUSSION

In an earlier experiment (Hively, 1962) a simple matching-to-sample program was empirically developed. In the present experiments, some properties of this program and the conditions under which children worked were systematically varied.

Although the program originally contained 120 items, 100 of these were found to be superfluous. The program was most efficient when it contained two slides in each step. Reducing the number of slides in each step to one destroyed the program's effectiveness, while increasing the number beyond two had little effect. Both the length of the program and the proportion of slides in each step affected the frequency of "post-learning errors".

No relationship was found between the sizes of the steps in the program and the optimum number of slides in each step. An "abridged" version of the program could not be improved by making it longer; abridging the program simply made the children's performances more variable.

Children's performances in the initial experiments (Hively, 1962) suggested that the program of discriminative stimulus change had a reinforcing effect. In the present experiments, this effect was found to be negligible.

A "run-to-criterion" contingency, which advanced the subject to a new step in the program whenever he responded correctly to a given number of consecutive slides, did not increase the program's efficiency. The size of the criterion needed to hold subjects in a given step when they began to make errors dictated a program which was unnecessarily long.

A "drive-back" contingency, which returned the subject to an earlier slide in the program each time he made an error, had no effect upon the performances of subjects working on the program with two slides in each step. The drive-back contingency did, however, substantially improve performance on the program with just one slide in each step.

A contingency requiring a short fixed-ratio of responses to the correct stimulus improved performance substantially. This contingency also furnishes a way of measuring response strength in a"multiple-choice" task.

In interpreting these data one should keep in mind the characteristics of this particular matching task. An additional group of children, comparable to those who participated in the present experiments, learned the Criterion Task in a short time when given verbal instructions rather than the program. In this sense, most of the children already "knew how to match": the program merely gave them their instructions. Some of the observed discontinuities may therefore be a function of the task. An investigation of programs designed to teach arbitrary associations rather than formal matching might reveal more continuities.

The dramatic effect of the fixed-ratio contingency upon the children's performances raises the embarrassing possibility that the preceding studies dealt mainly with parameters of an artifact, i.e., an imprecise contingency of reinforcement. However, the main effect of program length was still found under the fixed-ratio contingency and so were"post-learning errors".

A fixed, standard program with two slides in each step and the fixed-ratio contingency turned out to be the most effective of all the conditions studied. Subsequent experimentation has shown that it does not, however, bring all subjects under control: one can expect to lose control of about one subject in ten. I doubt that lengthening the program would make it more effective although I have not tried it with the fixed-ratio contingency.

The foregoing experiments underline the significance of Skinner's remark (Skinner, 1958) that learning curves reflect the properties of the apparatus more often than they reflect fundamental behavioral processes. However, they also suggest the possibility that one might eventually expose basic processes by attempting to arrange *optimum conditions* under which a given bit of behavior may be learned. One could argue that the changes in conditions which improved performance in the foregoing experiments were effective primarily because they removed some sources of uncontrolled variability in the training situation. (In this connection, it is worth noting that, insofar as they were effective, these conditions also wiped out "individual differences".)It seems reasonable to argue that if one could isolate the characteristics which made this matching-to-sample program effective, or if one could go somewhat further and discover some principles for designing an optimum program, one would then know a great deal about the basic processes involved.

## REFERENCES

Gellerman, L.W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, *42*, 207-208.

Hively, W. Programming stimuli in matching to sample. *J. exp. Anal. Behav.*, 1962, *5*, 279-298.

Hively, W. A multiple-choice visual discrimination apparatus. Harvard Teaching Machine Project, Final Report, 1964.

Skinner, B.F. Reinforcement today. *Amer. Psychol.*, 1958, *13*, 95.

## BASIC SKILLS: Specific Skills

5

PRECEDING PAGE BLANK-NOT FILMED

# SYSTEMATIC INSTRUCTION IN THE ATTAINMENT OF RIGHT-LEFT FORM CONCEPTS IN YOUNG AND RETARDED CHILDREN[1]

Sidney W. Bijou

## INTRODUCTION

In a review of the literature on concept attainment, Sigel ( 1964 ) points out that theories on the development of concepts may be divided into two major categories on the basis of whether or not such learning follows an invariant stage sequence. In other words, one group of theorists claims that the learning of concepts at a given level of complexity or difficulty requires a "readiness" or something similar to the setting conditions of the proper stage. Learning a higher-order concept requires previous acquisition of a lower-order concept of the same class. Instruction in conceptualization during the non-appropriate stage would have few or no beneficial consequences. Stages are believed to be generated by innate processes (*e.g.*, Gesell & Amatruda, 1941), or a combination of innate and experiential processes (*e.g.*, Tanner & Inhelder, 1960).

The other group of theorists believes that progress through invariant stages is not necessary and that advances in concept formation, like advances in other abilities and skills, come about through interactions between the individual and the environment. For learning to take place, however, it is necessary for the individual to have the equipment required to make appropriate responses (*e.g.*, say the appropriate words, or make the essential manual movements) and the environment to be so arranged that the critical classes of responses are strengthened.

Sigel also points out that the differences between the two points of view can be resolved by data from investigations based on longitudinal analysis or on an evaluation of the influence of teaching on concept learning. The latter consists of procedures which (1) evaluate the critical performance, (2) institute teaching procedures, and (3) re-evaluate the performance.

Investigations on the influence of teaching on conceptualization tend to conclude that instruction has little or no effect in accelerating concept attainment. Sigel expresses reservations about such a conclusion:

> In each of these studies the authors tried to teach the child some concept just before he was " ready " and failure was attributed to his immaturity and lack of experience. Possibly they did not employ sufficiently long training periods, *nor ascertain with any precision the necessary sequences as preludes to the learning.* (1964, p. 240)

This investigation is concerned with evaluating the influence of instruction on the development of nonverbal concepts involving rotated geometric forms with right-left orientational differences. The point of departure here is the application of programmed instruction techniques as the procedure for teaching. Giving instruction through programmed learning has several advantages from a research point of view. It permits the investigator (1) to describe to others in objective terms the essential characteristics of the materials and procedures and how they were developed, (2) to report in specific terms the effectiveness of the attempt, and (3) to point to other specific variations in stimulus material that have promise for the further improvement of the teaching procedure.

The study is reported in four main sections and an appendix. The first part describes the children on whom the material was developed and evaluated, the apparatus and procedures employed, and the stimulus material (the form program) used in training. The second section presents samples of the performance of children--two normal and two retarded. The third gives the overall findings; and the fourth, a discussion of results and potentialities of the approach.

The guide lines for the preparation of the initial stimulus materials, and the data which served as a basis for the successive revisions of the program, are given in the appendix. They are reported in detail here because it is believed that the procedures employed in the construction and evaluation of the "teaching materials" for this study could be of value to others interested in preparing training materials involving other types of concepts.

## METHOD AND PROCEDURE

Subjects.[2] The normal or non-retarded group, which was used to develop the program and provide data on its effectiveness on a normal group, consisted of 90 children from the preschool and primary grades of the Lesley-Ellis School, Cambridge, Massachusetts. There were 52 boys who ranged in age from 3 years 5 months to 6 years 11 months and 38 girls from 3 years 3 months to 6 years 8 months. Intelligence test scores were not available on these children. However, indirect bits of evidence, such as reports from teachers, educational and socio-economic status of parents, and an occasional sample of a child's test performance, indicated that the group was average to above average in intelligence.

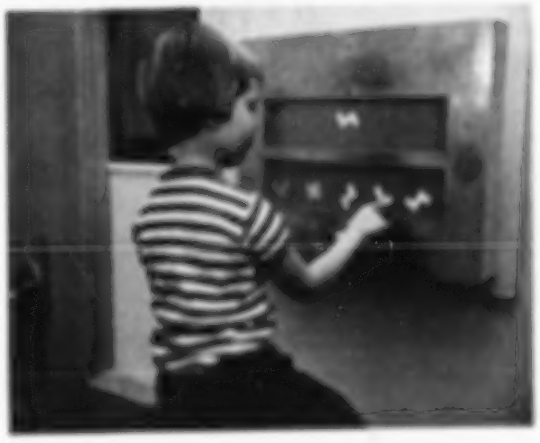


Fig. 1. Photograph of the apparatus.

The retarded group consisted of 89 children in residence at the Walter E. Fernald School, Waverly, Massachusetts. There were 75 boys who ranged from 6 years 4 months to 16 years 11 months in chronological age, from 3 years 10 months to 8 years 10 months in mental age test scores, and from 32 to 65 in IQ scores. The girls ranged from 12 years to 14 years 7 months in chronological age, from 4 years 11 months to 7 years 2 months in mental age test scores, and from 39 to 66 in IQ scores. All were classified as educable.

Apparatus. The apparatus, shown in Figure 1, is a variation on a device used extensively in infrahuman research. It was originally conceived by Skinner and further developed by Holland and Long to study " inductive reasoning " in young children ( Holland, 1960, 1963). A modification of it was used by Hively ( 1962, 1964 ) to study form discrimination in young children.

The device consists of a box with a panel of windows to display the stimuli as well as to register selections, a Bell and Howell 2 x 2 slide projector, an Esterline-Angus event recorder, and a circuitry rack. The *S* (subject) was seated so as to face the two horizontal, translucent, plastic windows. The upper window was a single unit; the lower one was divided into five equal parts. Geometric forms were projected on the windows from behind. One form appeared in the center of the large upper window and one in each of the five lower windows. The form in the upper window was designated as the sample, those in the lower windows as the matches. There was a temporal sequence. The sample was presented first. The matches appeared only after *S* pressed on the sample window. Temporal relationships between the sample and choices, order of slide presentations, and stimulus feed-back (light and chime) were controlled by switching-circuits.

The child indicated which of the five forms matched the sample by pressing on the window on which it appeared. If he made a correct selection a red light glowed momentarily, a chime sounded, the sample and the five choices disappeared, and another sample appeared. *S* could then produce the choices by pressing on the sample window, and proceed to make the next match. If he made an incorrect choice, the light and chime did not operate but the matches disappeared. (There was also an audible and clearly distinctive thud by the mechanism which blacked out the matches. ) Pressing on the sample window restored the choices (removed the blacking device ) and provided a second opportunity for matching. If the second response was also incorrect, the choices blacked out once more. Pressing again on the upper window restored the matches again. This sequence ( press on sample window to display matches; selection of a mismatch; disappearance of matches; press on sample window to display matches, etc. ) would be continued until the correct response was made. Making the correct response removed both the sample and the matches and presented another sample. Under this arrangement of contingencies the last response to a problem was always the correct response.

However, when a correct response was preceded by one or more incorrect responses, the next slide presented was the previous slide. That is to say, the next slide after an error was a repeat of the slide previously " passed " rather than an advancement to the next new slide in the sequence. Thus, a child could not move forward to the next task in the program after an error without first reacting a second time to both the preceding slide just "passed" and the slide just "missed." It was therefore possible for a *S* to make "new" errors on slides previously passed without error, and to go backward rather than forward in the program.

*Instructions.* *S* was brought to the experimental room by a young lady experimenter.[3] After inviting him to sit in the chair and after adjusting the height of the chair so that his eyes were approximately on a level with the strip separating the upper and lower windows, the experimenter sat beside him and said: "Put your finger on this (sample in upper window) and press it." "Find one here (choices in lower window ) just like it and press it." "Good."

After three correct consecutive selections each followed by "Good," the experimenter said: "Try to get it right the first time," and went to the rear of the room where she monitored the apparatus and observed the child's behavior through a mirror. At the end of each tray (usually 40 slides) she said: "Very good. I have more for you." When the next tray was in operational position she said: "Go ahead."

Stimulus Material. The stimuli consisted of 10 forms cut from 1 1/2-inch squares of red, yellow, green, and blue paper.[4] Each was composed of five 1/2-inch squares; three formed a vertical column and two served as appendages protruding from the slides. In some forms the two squares were attached to the same side of the column (e.g., L ); in others they were on opposite sides (e.g., + ). Three of the forms (P, ʃ , and ɟ ) had mirror-image possibilities (¶, ʅ, and ɬ ). Thus in each form, three-fifths of the total area was oriented north and south (a stack of three squares forming the vertical column); and two-fifths, east and west ( two squares forming the appendages). The 10 forms used throughout the study are shown in Column 1, Figure 2.

The sample and choices constituting a frame were photographed on 16 mm. color-sensitive film and developed into standard 2x2 slides. Two forms of the program were constructed, one with one color in each slide (sometimes red, yellow, blue, or green) and the other with four colors mixed within a slide.

The program was made up of 270 slides divided into Elementary, Intermediate, and Advanced Sets. The guidelines for constructing the initial slides and the trial runs on children from the Lesley-Ellis School and subsequent revisions are given in the appendix.

Elementary Set--24 Slides. The Elementary Set, which evolved from four major revisions, started with: a solid colored circle, square, and triangle constructed from 1 1/2-inch squares. These simple forms served to ease the youngster into operating the apparatus. (Press on sample in upper window to produce choices. Press on lower window to indicate selection and to produce the next sample. Repeat.) They also served to instruct the child to engage in the attending behavior necessary for making the matching response as described in the next paragraph.

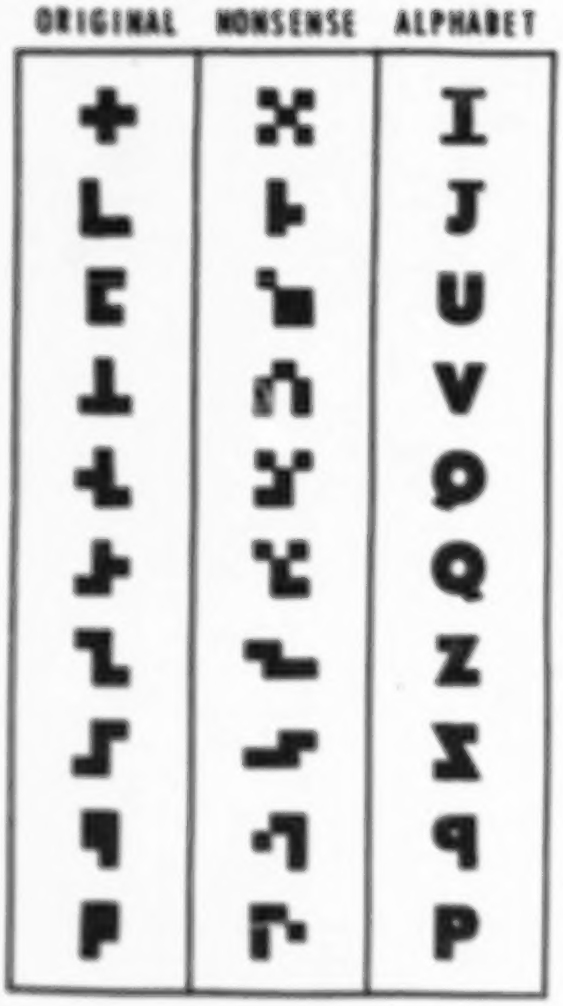


Fig. 2. Forms used in mirror-image training and pre- and post-testing.

In the first frame the red ball was the sample; and the red ball, square, and triangle were presented as choices in the first, third, and fifth windows from the S's left to right (see Figure 3, slide 1-1). In the second and third slides, three choices were also presented, but in different windows, and the square and then the triangle replaced the circle as sample. The

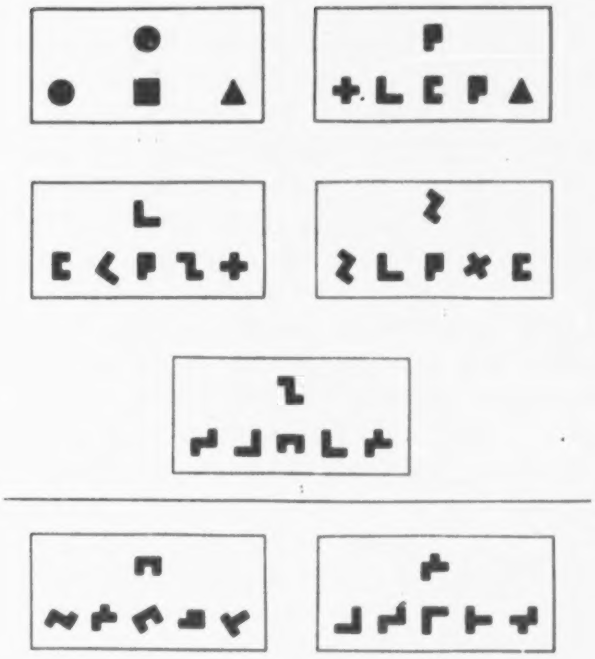


Fig. 3. Schematic drawings of sample slides from the Elementary (above the horizontal line) and Intermediate Sets.

first experimental form ( [ ) was added in the fourth frame, and the second(+) was added in the sixth. Hence, from the sixth slide on, five choices were presented. Three additional forms (P, L, ⅂) gradually replaced the ball, square, and triangle, and two others ( T , ⊣)were made to alternate with the first five,making a total of seven different experimental forms. The ball last appeared in frame seven, the square in frame eight, and the triangle in ten. Slide ten is shown in Figure 3, 1-10.

In slide 13 the correct match appeared in rotated positions for the first time. (See Figure 3, 1-13.) The sample was in the upright position, and the correct match was rotated 30 degrees to the east. Rotations to the east of vertical were referred to as plus; to the west, as minus.

In slide 14 ( Figure 3, 1-14 ), the sample and correct match were rotated plus 30 degrees and one of the incorrect matches (+) was at minus 30 degrees. It will be noted that although both the sample and the match were rotated in slide 14, the discrepancy between them in rotational degrees was zero.

From slides 15 to 24, the end of the Elementary Set(Figure 3,1-24), the sample was at zero, and the matches gradually varied to plus or minus 90 degrees, thereby presenting rotational discrepancies between samples and matches up to 90 degrees. Except for slide 14, the sample was always in the upright position in the Elementary Set.

Intermediate Set--40 Slides. The Intermediate Set, like the Elementary, evolved from four major revisions. This sequence was designed to train in discriminations when both the samples and the matches are rotated in the vertical plane through plus and minus 180 degrees and when rotational discrepancies between the sample and the matches vary from zero to 180 degrees.

The first four slides in the Intermediate Set presented the sample at zero, while the matching stimuli varied plus or minus 30, 60, or 90 degrees. In the other 36 slides, the samples were gradually rotated,as shown below, while the matches continued to vary plus or minus 30, 60, or 90 degrees.

| Slides | Rotation of Sample in Degrees |
|---|---|
| 1-5 to 1-12 | 30 and 60 |
| 1-13 to 1-36 | 30, 60, and 90 |
| 1-37 to 1-40 | 90, 150, and 180 |

The discrepancies between the rotational positions of the samples and correct matches ranged from 60 to 180 degrees. Two slides in the Intermediate Set are shown below the horizontal line in Figure 3. Slide 1-20 (the middle slide in the set) presents the sample at plus 90 degrees and the correct match at plus 60 degrees, with a rotational discrepancy between them of 30 degrees. Slide 1-40 (the last slide in the set ) shows the sample at plus 90 degrees and the correct choice at minus 90 degrees, with a rotational discrepancy between them of 180 degrees.

Advanced Set--206 Slides. The Advanced Set, which developed from five major revisions, was designed to train in discriminating between mirror-images and non-mirror-images of three forms, presented in upright positions and then with rotations in the vertical plane. In this series, only the P, ʃ, and ┣ were samples. Half the time they were presented oriented in the opposite direction, i.e., as ꟼ, ⅂, and ┫. Because of its length, the Advanced Set was divided into five units corresponding to five trays of slides.

Unit 1 (slides 1-1 to 1-46). To introduce in small stimulus units the mirror-image of the sample as an incorrect stimulus, this sequence began with 15 specially prepared slides. In these ( 1 ) the mirror-image of ꟼ was gradually evolved by raising, in successive slides, the lower horizontal square of [, ( 2 ) the mirror-image of ⅂ was gradually developed by lowering the square to the west of the vertical column ( "the wing" ) of ʃ, and (3) the mirror-image of ┫ was gradually formed by lowering the square to the west of the vertical column (the left arm) of ┼. In other words, changes were made in three incorrect non-mirror-image forms to make them into incorrect mirror-images with respect to forms P, ⅂, and ┣ when they were presented as samples. This

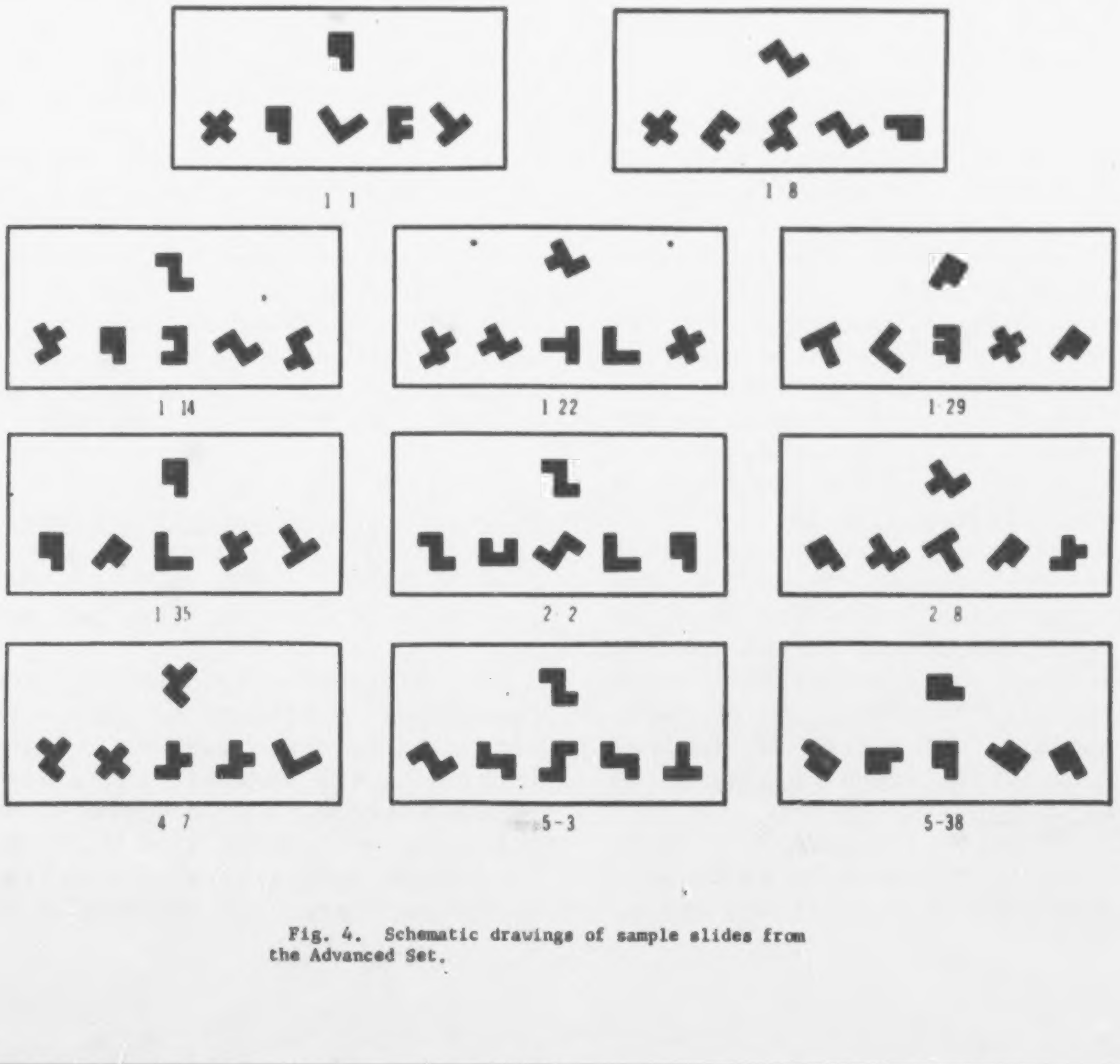


Fig. 4. Schematic drawings of sample slides from the Advanced Set.

"fading-in" process was accomplished in six stages.

Stage 1, slides 1-1 to 1-6. In these six slides the three key forms served as samples twice, once oriented to the left ( ᑫ , ᘔ , and ⊥ ) and once toward the right ( ᑭ , ʃ , and ɟ ). Vertical plane rotations of the sample and the correct stimulus were either zero or plus or minus 30 degrees, with the rotational discrepancies between sample and correct stimulus at zero.

When the ᑫ was a sample, one of the incorrect matches was a [ with the lower right square elevated on the vertical column one-half the distance of a square. Slide 1-1 in Figure 4 shows the altered figure in the fourth position. When ᘔ served as sample, one incorrect choice was a figure with the left "wing" of ⊤ lowered one-half the distance of a square; and when ⊥ was sample, one of the wrong choices was a figure with the left "arm" of + lowered one-half the distance of a square.

Stage 2, slides 1-7 to 1-12. In these six frames the incorrect potential mirror-images of the ᑫ , ᘔ , and ⊥ had the critical squares moved up or down the vertical column an additional one-fourth of the distance of a square. See slide 1-8 in Figure 4. Each of the three forms was presented as sample twice, once oriented toward the right and once toward the left. Vertical plane rotations of the sample and correct stimulus were zero or plus or minus 30 degrees, with zero rotational discrepancies between them.

Stage 3, slides 1-13 to 1-18. In these six slides the incorrect potential mirror-images of ᑫ , ᘔ , and ⊥ had the critical squares moved up or down an additional one-eighth of the distance of a square. See the form in the fifth position in slide 1-14, Figure 4. And as in the other sequences, each form was presented an equal number of times oriented toward the right and left, and rotations of the sample and correct match were zero or plus or minus 30 degrees, with rotational discrepancies between them no greater than 30 degrees.

Stage 4, slides 1-19 to 1-28. In these 10 slides the incorrect potential mirror-image forms of the "flag," "Z" figure, and winged figure had the critical squares moved again, this time one-sixteenth of the distance of a square. See the form in the first position in slide 1-22, Figure 4. The flag and "Z" figures were samples in six slides, and the winged figure was the sample in four. As before, they appeared in opposite orientations, and rotations of the sample and correct stimulus were zero or plus or minus 30 degrees, with rotational discrepancies of zero.

Stage 5, slides 1-29 to 1-34. In these six slides the incorrect potential mirror-image forms of the three figures used as samples had the critical squares moved once more, this time one-thirty-second of the distance of a square. As an example, see the form in the third position in slide 1-29, Figure 4. Each of the forms was presented as samples twice, once oriented to the right and once to the left. Rotations of the sample and correct stimulus were zero or plus or minus 30 degrees, with four discrepancies at zero and two at 30 degrees.

Stage 6, slides 1-35 to 1-46. In these 12 slides, full mirror-images appeared as incorrect matches for the three forms. The first one, with the flag oriented toward the left as sample, is shown in Figure 4, 1-35. Although the samples and matches in this series were either at zero or at plus or minus 30 degrees, the discrepancy between samples and correct matches was zero, except for slide 1-40, which was at 30 degrees.

Unit 2 (slides 2-1 to 2-40.) The 40 slides in unit 2 were designed to give further training in discriminating a mirror-image from a non-mirror-image of the sample in the context of: (1) increased vertical plane rotations of the sample and matches, (2) increased discrepancies between sample and correct match, (3) increased discrepancies between sample and incorrect match, in the form of a mirror-image of the sample, and (4) increased discrepancies between correct match and incorrect mirror-image match. The rotational sequences were as follows:

1. Samples were rotated zero, 30, 60, and 90 degrees.
2. Correct matches were rotated zero, 30, and 60 degrees.
3. Incorrect mirror-image matches were rotated zero, 30, 60, and 90 degrees.

Discrepancies between sample and correct matches ranged from zero to plus or minus 60 degrees--25 slides with zero, 11 with 30, and 4 with 60--and new rotational positions of incorrect mirror-images (declinations from the vertical position in 30 - degree units) were introduced in windows adjacent to correct matches, and then gradually separated by one, two, and three windows. In Figure 4, slide 2-2 shows a separation of one window between correct and incorrect mirror-image match, and slide 2-8 shows a separation of two windows

Unit 3 (slides 3-1 to 3-40). The slides in unit 3 were designed to extend training in discrimination of the correct stimulus from its mirror-image by increasing the number of items with discrepancies at 30, 60, and 90 degrees. Because of the high error rate associated with 90-degree discrepancies, the step between 60 and 90 degrees was divided in half. This was accomplished with 6 slides with 75-degree rotations. In this unit, samples, correct matches, and incorrect mirror-image matches were rotated zero, 30, 60, 75, 90, 105, and 135 degrees.

Discrepancies between sample and correct matches were as follows:

| Slides | Rotational Differences in Degrees |
|---|---|
| 3-1 to 3-14 | 30 (with one exception at 60) |
| 3-15 to 3-25 | 60 (with one exception at 90) |
| 3-26 to 3-31 | 75 |
| 3-32 to 3-40 | 90 |

The correct stimulus and incorrect mirror-image were gradually separated in terms of window location as follows:

| Slides | Window Location |
|---|---|
| 3-1 to 3-7 | Adjacent |
| 3-8 to 3-22 | One window between |
| 3-23 to 3-34 | Two windows between |
| 3-35 to 3-40 | One window between (match and sample at 90 or zero degrees) |

Unit 4 ( slides 4-1 to 4-40 ). Starting with slide 4-7 in unit 4, two incorrect mirror-images of the sample appeared. Rotational deviations between sample and correct match ranged from 0 to 90 degrees. In addition to the practice of introducing new rotational positions of incorrect mirror-images in windows adjacent to correct matches and gradually presenting them in most distant windows, the two incorrect mirror-images appeared initially in parallel orientation; that is, both were at zero, 30, or 60 degrees. (Slide 4-7 in Figure 4 shows parallel incorrect mirror-images of the winged form. ) The two incorrect mirror-images then appeared in different rotational positions with respect to each other, the correct match, and the sample. The sequence in unit 4 was as follows:

1. One incorrect mirror-image among the matches

| Slides | Rotational Differences in Degrees |
|---|---|
| 4-1 to 4-2 | 60 |
| 4-3 to 4-6 | 90 |

2. Two incorrect mirror-images among the matches

| | |
|---|---|
| 4-7 to 4-13 | 0 |
| 4-14 to 4-20 | 30 |
| 4-21 to 4-31 | 60 |
| 4-32 to 4-40 | 90 |

Slide 4-7 is shown in Figure 4.

Unit 5 (slides 5-1 to 5-40). Unit 5 presented matching problems with three and four incorrect mirror-images. Rotations appeared in gradual stages, and incorrect mirror-images were displayed in parallel deviations in the initial presentations and then varied in rotational positions independently. The order of slides was as follows:

1. Two incorrect mirror-images among the matches

| Slides | Rotational Differences in Degrees |
|---|---|
| 5-1 | 120 |

2. Three incorrect mirror-images among the matches

| | |
|---|---|
| 5-2 to 5-5 | 30 |
| 5-6 to 5-10 | 60 |
| 5-11 to 5-19 | 90 |

3. Four incorrect mirror-images among the matches

| | |
|---|---|
| 5-20 to 5-23 | 30 |
| 5-24 to 5-29 | 60 |
| 5-30 to 5-37 | 90 |
| 5-38 to 5-40 | 150 |

In Figure 4, slide 5-3 shows a problem with three incorrect mirror-images and rotations to 90 degrees, and slide 5-38 shows one with four incorrect mirror-images and rotations to 120 degrees.

Pre- and Post-Tests. Three sets of slides, 20 in each, were prepared to evaluate training from the program. One set, designated as "Original," was made up of duplicates taken directly from the program -- 4 from the Intermediate and 16 from the Advanced Set. A second set, called the "New Nonsense" test was specially prepared. These forms were like those in the Original test in area and color but differed in shape. Performances in relation to these forms were to provide data on stimulus control of material similar to that in the program. ( The relative discriminability of the shapes in the Original and New Nonsense was unknown and hence complicated interpretation of differences in performance on the two tests. ) The third test, the "Alphabet," was composed of ten letters drawn at random. The area and color of the letters were also the same as those in the Original. Performance on this series would provide some information on stimulus control of letters similar in construction to the forms in the program.

Figure 2 shows the forms in the three tests. They are arranged in three columns with " equivalent " forms in the rows. For example, a slide in the Original test involving the first five forms in column 1 had a corresponding slide in the New Nonsense test made up of the first five forms in column 2, and a corresponding slide in the Alphabet test made up of the first five letters in column 3.

## SAMPLE DATA ON INDIVIDUAL CHILDREN

The data, recorded by Esterline-Angus event markers, consisted of the time between pressing on the sample window and the match window, correct or incorrect. This information is transcribed into graphic form as shown in Figure 5. Figure 5 shows the performance of a Lesley-Ellis girl (S64), 4 years 5 months of age, on the Elementary, Intermediate, and Advanced Sets, which were administered in four sessions. Time is plotted in 20-second units on the abscissa;and the order number of slides, on the ordinate. The cumulative curves are constructed by representing a correct response as a step up to the next slide and duration of response time as the length of the tread of the step. When the slide order was reversed, following an incorrect match,the curve shows a regression to the previous step. Errors are represented as vertical strokes, twice the size of a step.

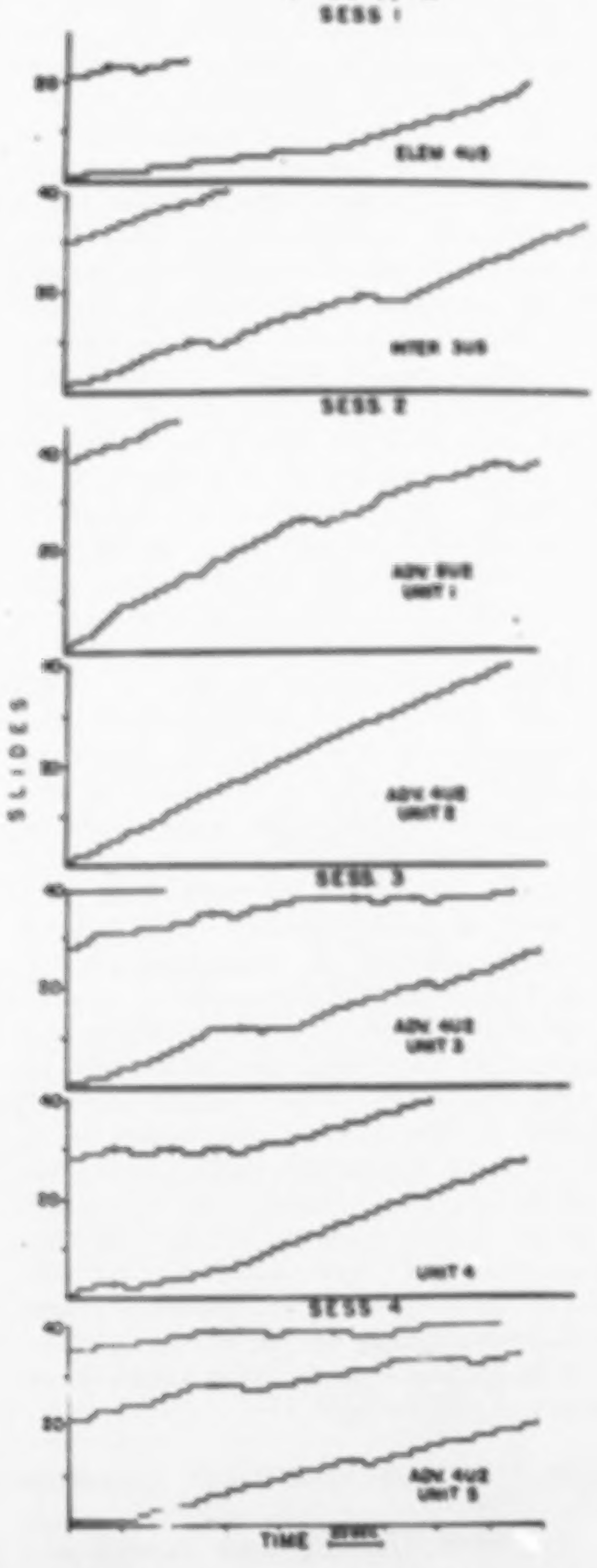


Fig. 5. Performance of a girl, S64, 4 years 5 months, on the Elementary, Intermediate, and Advanced Sets administered in four sessions.

Figure 5 may be described as follows: In session 1, first and second graphs from the top, the girl's matching behavior was rather slow and deliberate on the first seven frames of the Elementary series (4U5). These were the slides with the ball, solid square, and triangle. She made one error near the end of the set on slide 1-23, which involved rotational discrepancies at 60 degrees. The time for making the incorrect response, shown by the vertical stroke, was much shorter than the time for the correct response. Regression in the curve represents a return to slide 1-22. (The curve is brought back to the ordinate line at slide 1-21 to avoid extending the graph laterally.) Her performance on the Intermediate Set (3U5) shows a fairly constant rate of matching and an error on slide 1-10 and another on 1-19. The curve is brought back to the ordinate on slide 1-30 for the purpose of convenience.

In session 2, third and fourth graphs from the top, units 1 and 2 of the Advanced Set (5U2) were administered. On unit 1, which presented partial and then full mirror images of the sample, she showed a deceleration of rate after an error on slide 1-26 ( fourth stage construction of mirror-image). The second error was made on slide 1-37, which was the third slide to appear with a full mirror-image of the sample. On unit 2, which had one incorrect mirror - image match and small rotational discrepancies (30 and 60 degrees), she was fairly fast and performed without error.

In session 3, second and third graphs from the bottom, her matching rate was slower in unit 3 than it had been in unit 2, and she made five errors. Although only one incorrect mirror-image match was involved in these frames, rotational discrepancies ranged as high as 135 degrees. Attention is called to the last two errors. She made both on slide 3-37, the second a repeat of the first after a correct response to slide 3-36. On unit 4, which involved two mirror-images, she made four errors, three on slide 4-30.

In session 4, bottom graph, her rate of matching progressively decreased on unit 5, which involved three and then four incorrect mirror-image matches. It should be noted that four of her seven errors were "successive errors." That is, two were made on slide 5-28 prior to making the correct match and two on slide 5-39 under the same circumstances.

A second example is shown in Figure 6. This S (S12), a Lesley-Ellis boy, five years eleven months of age, was started on the Intermediate Set (1U5). As shown in the upper left pair of graphs, his rate was slow and fairly constant in unit 1, but increased in unit 2. Only one error was made in unit 2. In the same session he was given units 1 and 2 of the Advanced Set (4U5). As shown in the lower left pair of graphs, his rate steadily decreased and he made four errors on unit 1 and one on unit 2.

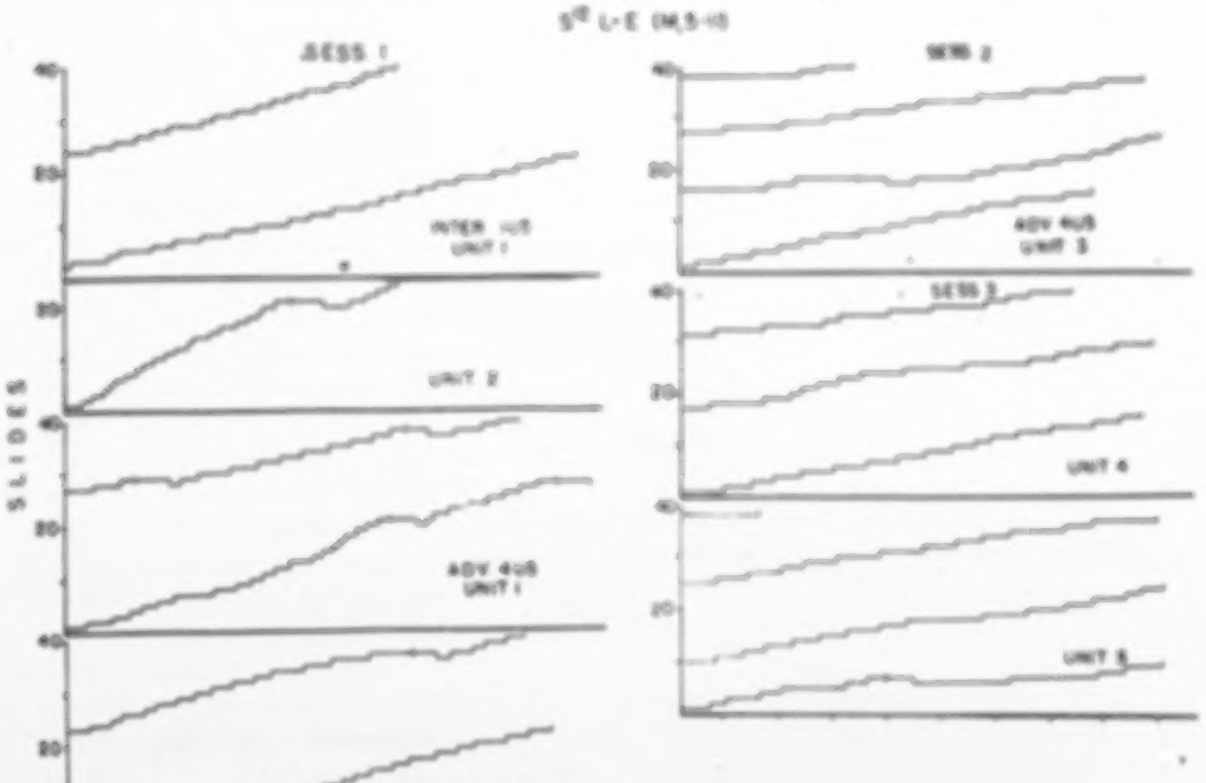


Fig. 6. Performance of a boy, S12, 5 years 11 months, on the Intermediate and Advanced Sets administered in four sessions.

In Session 2, he was given only unit 3 of the Advanced Set. He continued to increase his time and made only one error, on slide 3-18.

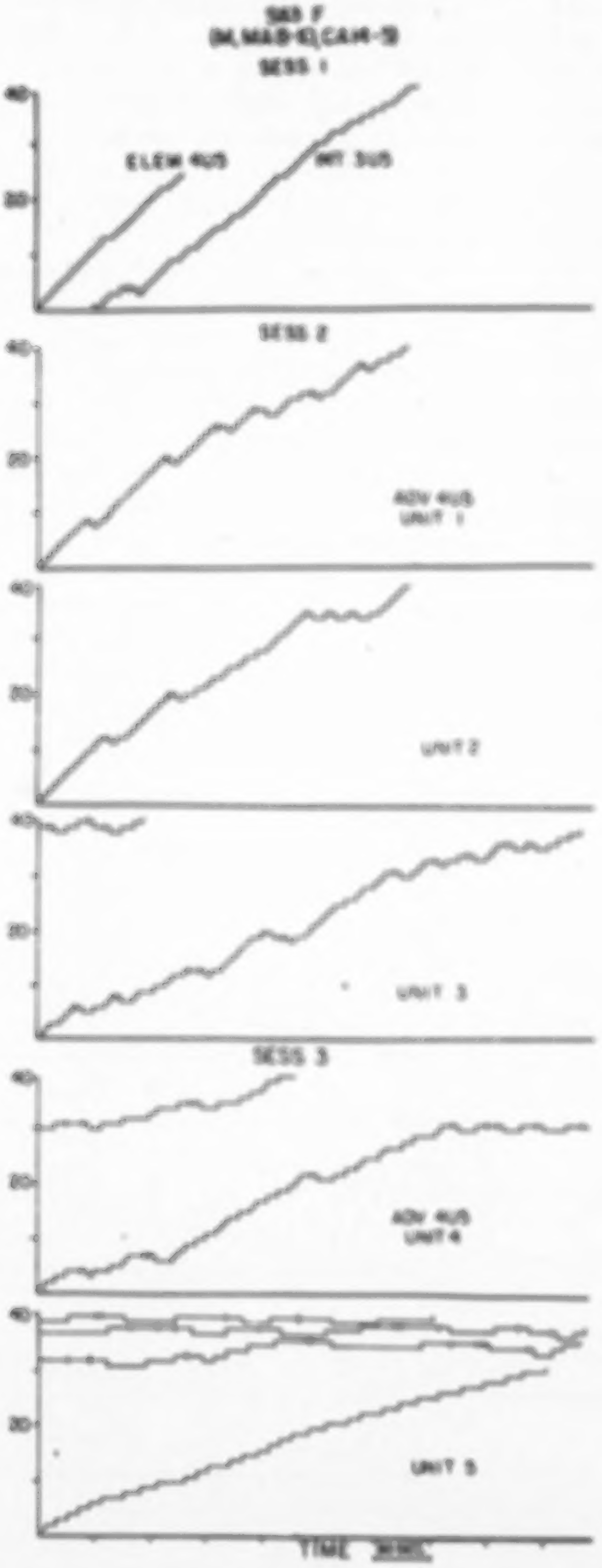


Fig. 7. Performance of a boy, §63F, 14 years old, on the Elementary, Intermediate, and Advanced Sets administered in three sessions.

In session 3, units 4 and 5 were given in the same series. His matching continued at a low rate and he made one error on unit 5, on slide 5-7.

A third example is shown in Figure 7. This is the performance of a Fernald boy ( §63 ), 14 years of age, with a mental age of 8 years 10 months. In session 1, he went through the Elementary and Intermediate Sets at a rather high rate of matching. One first error was made on slide 4 of the Intermediate sequence. In session 2, his high rate was maintained on unit 1 of the Advanced Set but he made six first errors. On unit 2, he made five errors, three first errors and two repetitive errors on slide 2-35. His rate of matching decreased slightly after frame 2-20, the slide on which he made his second error. On unit 3 he made a total of 14 errors. For the first time he made two errors before making the correct match (slide 3-13 shows two pips preceding a regression in the curve ), and two errors on slides previously passed on the first attempt ( slides 3-19 and 3-39 ). In session 3, he made 13 errors on unit 4, and his rate of matching decreased further. In four instances ( slides 4-7, 4-30, 4-31, and 4-35) he made two errors before making the correct selection. In one situation he alternated between slide 4-30 and slide 4-31 four times, repeating the same errors. On unit 5 his rate of matching steadily decreased over the first 31 slides and he performed without error. After slide 5-31 he made 21 errors.

A fourth example, shown in Figure 8, is the record of an 11-year-old Fernald boy (§6) with a mental age of 6-0. He was started on the Intermediate Set and went through units 1 and 2 at a high rate without error. He also completed units 1 and 2 of the Advanced Set, making

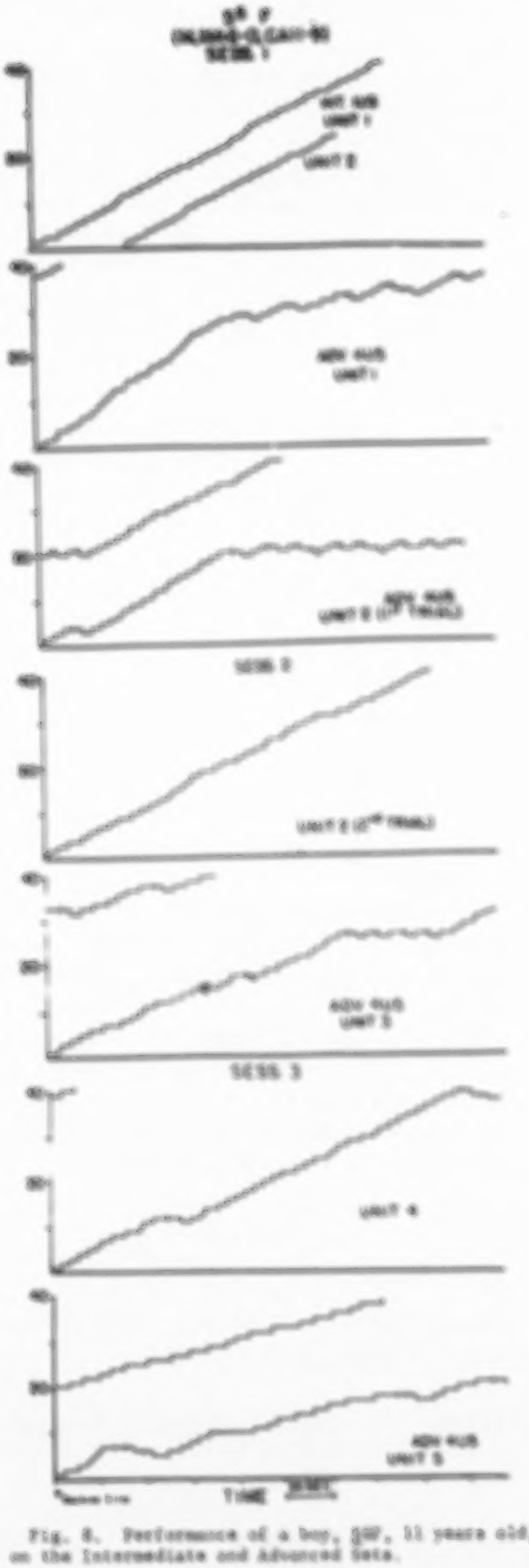


Fig. 8. Performance of a boy, GDF, 11 years old, on the Intermediate and Advanced Sets.

6 errors on the first sequence and 13 errors on the second. In session 2 the second unit was re-presented and he went through it this time without error. In this session, unit 3 was also given, and he made 7 errors on it. (The circle with one dot indicates an apparatus "error.") Finally in session 3 he completed unit 4 with only 3 errors and unit 5 with 6.

## RESULTS

The Fernald and Lesley - Ellis children will be compared in terms of their performances on the ( 1 ) pre - training and post - training tests, (2) Elementary and Intermediate Sets(stimuli without mirror-image forms), and ( 3 ) the mixed-color and single-color series. In addition, data will be presented on the Fernald children with respect to performance on the Advanced Set and the relationship between intelligence test scores and progress in the program.

### Comparisons of Fernald and Lesley-Ellis Children

Pre-training and post-training tests. The effect of training on the program was evaluated by three sets of slides from forms shown in Figure 2. One test, designated as "Original," was made up of duplicates from the program;4 were from the Intermediate Set, and 16 were from the Advanced.The second test, called the "New Nonsense" test contained forms similar to the "Original" in area and color but different in shape. ( See the forms in the middle column of Figure 2. ) The third, the " Alphabet " test, was made up of 10 capital letters drawn at random. The letters were prepared so as to be comparable in size, area, and color to the forms in the other two sets. ( See the forms in the right-hand column of Figure 2. ) There were 20 slides in each test series.

The tests were administered prior to training as follows: ( 1 ) The usual instructions were given and

*S* matched the first 15 slides in the Elementary Set to give him an opportunity to learn how to perform the task. ( 2 ) The instructions were repeated and *S* was given the 60 slides in the three tests. The light and chime following correct matches, and the slide back-up action of the projector following incorrect responses, were not used. (3) The instructions were given again and *S* was started on slide 1-16 of the Elementary Set with light and chime following correct responses, and slide back-up action of the projector following incorrect responses. (To evaluate the influence of the pre-tests on performance in the program, some *S*s started the training series with slide 1-1.)

Table I shows the mean error frequencies for 8 Fernald and 5 Lesley-Ellis children on the tests *before* training on the form program. All errors are first errors, since each response, correct or incorrect, moved the program forward. As will be noted, the mental age scores of Fernald children ranged about a year higher than the chronological age of Lesley-Ellis children.

Both groups made errors on about 50% of the slides. The Fernald children, however, made somewhat fewer errors on the Alphabet series (the upper figure in the right-hand column). The chances are that the Fernald children had more extensive histories with letters, words, and sentences.

Mean errors on the test administered to 10 Fernald and 6 Lesley-Ellis children *after* training on the program are shown in Table II. Like the *S*s shown in Table I, the Fernald children had mental age scores ranging above the chronological ages of the Lesley-Ellis children. In this instance the Fernald children were two years higher.

The 10 Fernald children are presented in two groups of 5 each on the basis of the number of days between completion of the program and administration of the post-tests. The first 5 *S*s were tested within a period of 28 days, the second 5 between 30 and 59 days. The Lesley-Ellis children were also divided into two groups on the same basis. The first group of 5 were tested within 16 days after finishing the program, and one *S* was tested after 37 days had elapsed.

A comparison of the error frequencies in Tables I and II makes three relationships apparent. First, Fernald and Lesley-Ellis children made fewer errors on the tests after training and were not strikingly different on the amount of improvement made. Second, both groups performed similarly on the

Table I

Mean errors of Fernald and Lesley-Ellis children on the Original, New Nonsense, and Alphabet tests *before* training

| | N | CA | MA | Original (20 slides) | New Nonsense (20 slides) | Alphabet (20 slides) |
|---|---|---|---|---|---|---|
| Fernald | 8 | 11-10 to 14-7 | 4-11 to 7-2 | 10.3 | 10.6 | 8.9 |
| Lesley-Ellis | 5 | 4-9 to 6-3 | | 11.6 | 11.6 | 10.6 |

### Table II

Mean errors of Fernald and Lesley-Ellis children on the Original, New Nonsense, and Alphabet tests *after* training

| | N | CA | MA | Interval (days) | Original (20 slides) | New Nonsense (20 slides) | Alphabet (20 slides) |
|---|---|---|---|---|---|---|---|
| Fernald | 5 | 11-3 to 16-0 | 6-0 to 8-10 | 0-28 | 3.4 | 4.4 | 2.2 |
| | 5 | | | 30-59 | 3.6 | 7.0 | 2.0 |
| Total | | | | | 3.5 | 5.7 | 2.1 |
| Lesley-Ellis | 5 | 5-10 to 7-0 | | 0-16 | 2.0 | 3.6 | 2.2 |
| | 1 | | | 37 | 3.0 | 8.0 | 1.0 |
| Total | | | | | 2.2 | 4.3 | 2.0 |

### Table III

Mean errors of Fernald and Lesley-Ellis children on three tests (60 slides) before and after training on the basis of window location

| | N | CA | MA | Window Location* (Left to right) | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | 1 | 2 | 3 | 4 | 5 |
| **Before Training** | | | | | | | | |
| Fernald | 8 | 11-10 to 14-7 | 4-11 to 7-2 | 2.3 | 5.1 | 14.6 | 5.4 | 2.8 |
| Lesley-Ellis | 5 | 4-9 to 6-3 | | 2.6 | 5.0 | 19.8 | 4.4 | 2.8 |
| **After Training** | | | | | | | | |
| Fernald | 10 | 11-3 to 16-0 | 6-0 to 8-10 | 1.6 | 2.1 | 2.4 | 3.3 | 1.9 |
| Lesley-Ellis | 6 | 5-10 to 7-0 | | 2.0 | 2.0 | 2.0 | 1.7 | 0.8 |

---

*#1 is at the *S*'s extreme left and #5 at his extreme right.

post - tests, with most errors on the New Nonsense series and least on the Alphabet. Third, aside from the slightly fewer errors by Fernald children on the Alphabet series before training, both groups performed similarly.

Individual data on pre- and post-tests are provided by a 5 year 10 month old Lesley -Ellis girl (S90) who showed a pattern of errors similar to the group data shown in Tables I and II. For her, three days elapsed between pre- and post-testing, and the post-tests were given immediately after completion of the program. Her distribution of errors is as follows:

| | Errors | |
|---|---|---|
| | Pre-tests | Post-tests |
| Original | 10 | 3 |
| New Nonsense | 11 | 3 |
| Alphabet | 10 | 3 |

Errors on pre- and post - tests may also be analyzed on the basis of their occurrences in the windows of the matches. Information of this sort bears on the possible influence of relative position of the matches in a row. It is to be remembered that the windows were referred to by numbers from one to five, from the subject's left to right, and that the sample was always in the upper window above window 3.

The upper part of Table III shows the performances on the tests before training in terms of mean errors for each window for 8 Fernald and 5 Lesley-Ellis children. For both groups the highest number of errors was made on window 3, the next highest on windows 2 and 4, and the lowest on 1 and 5, a kind of normal distribution curve.

The lower part of Table III shows mean errors on the tests by window for 10 Fernald and 6 Lesley-Ellis children after training. For both groups, the number of errors over windows tends to be distributed evenly. Comparison of the upper and lower sections of the table shows that in addition to the differences in distribution, the means for the post-tests are lower than for the pre-tests.

Individual data on pre- and post-tests provided by the Lesley-Ellis girl (S90) cited above show error frequencies and distributions similar to the group data in Table III.

| Window | Errors | |
|---|---|---|
| | Pre-tests | Post-tests |
| 1 | 2 | 0 |
| 2 | 6 | 4 |
| 3 | 21 | 2 |
| 4 | 2 | 2 |
| 5 | 0 | 1 |

Elementary and Intermediate Sets. Table IV shows the number of errors by 27 Fernald boys on the

Table IV

Errors of Fernald boys on 24 slides of the Elementary Set (405)

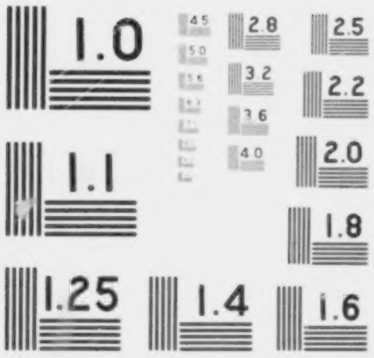
(N=27)

| | | | Errors | | |
|---|---|---|---|---|---|
| S | MA | CA | First | Repetitive | Outcome |
| S45 | 4-4 | 6-10 | 1 | 0 | Finished |
| S43 | 4-9 | 7-5 | 10 | 11 | Finished |
| S40 | 5-8 | 14-4 | 2 | 0 | Finished |
| S60 | 6-2 | 15-8 | 0 | 0 | Finished |
| S64 | 6-8 | 11-10 | 0 | 0 | Finished |
| S67 | 6-8 | 14-9 | 0 | 0 | Finished |
| S42 | 6-9 | 9-11 | 0 | 0 | Finished |
| S65 | 6-10 | 13-10 | 1 | 0 | Finished |
| S37 | 6-10 | 16-11 | 0 | 0 | Finished |
| S69 | 7-0 | 11-11 | 3 | 0 | Finished |
| S66 | 7-0 | 15-7 | 0 | 0 | Finished |
| S70 | 7-2 | 13-3 | 0 | 0 | Finished |
| S39 | 7-4 | 12-10 | 0 | 0 | Finished |
| S72 | 7-4 | 13-8 | 1 | 0 | Finished |
| S62 | 7-4 | 14-7 | 0 | 0 | Finished |
| S71 | 7-4 | 13-5 | 0 | 0 | Finished |
| S68 | 7-4 | 12-7 | 0 | 0 | Finished |
| S52 | 7-10 | 11-5 | 1 | 0 | Finished |
| S36 | 7-10 | 14-5 | 0 | 0 | Finished |
| S57 | 8-2 | 13-5 | 1 | 0 | Finished |
| S38 | 8-2 | 15-6 | 1 | 0 | Finished |
| S41 | 8-4 | 14-7 | 0 | 0 | Finished |
| S73 | 8-8 | 13-3 | 0 | 0 | Finished |
| S75 | 8-8 | 13-9 | 0 | 0 | Finished |
| S61 | 8-8 | 15-11 | 0 | 0 | Finished |
| S59 | 8-8 | 16-0 | 0 | 0 | Finished |
| S63 | 8-10 | 14-5 | 0 | 0 | Finished |

3 OF 4

ED

003552


1.0
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6

MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963

Elementary Set (4U5) of 24 slides. The Ss ranged in age from 6-10 to 14-5 and had intelligence test mental age scores from 4-4 to 8-10. All but S 43 made relatively few errors, and all finished the sequence. S43 made 10 first errors and 11 repetitive errors. He made several errors on the second and third slides; so the instructions were repeated. After that, whenever he made an error he responded to several other windows before making the correct match. Despite his 10 first errors and 11 repetitive errors, he was permitted to continue since he was making progress.

Table V shows the errors for Lesley-Ellis children on the same material. The 9 Ss ranged in age from 4-0 to 6-0. (As pointed out, intelligence test mental age scores were not available for these children.) This group made only three first errors.

To summarize: the performances of Fernald and Lesley-Ellis children on the Elementary Set, which involved simple rotations in the vertical plane, were similar except for one Fernald child (S43) who made 11 errors. This statement would hold even if the comparison were limited to the nine Fernald children with mental age intelligence test scores between 4-4 and 6-10.

Table V

Errors of Lesley-Ellis children on 24 slides of the Elementary Set (4U5)

(N-9)

| S | CA | Sex | Errors First | Errors Repetitive | Outcome |
|---|---|---|---|---|---|
| S54 | 4-0 | M | 0 | 0 | Finished |
| S55 | 4-3 | M | 3 | 0 | Finished |
| S64 | 4-5 | M | 1 | 0 | Finished |
| S70 | 4-6 | F | 0 | 0 | Finished |
| S52 | 4-10 | F | 0 | 0 | Finished |
| S60 | 5-8 | F | 0 | 0 | Finished |
| S61 | 5-9 | F | 0 | 0 | Finished |
| S67 | 6-0 | F | 1 | 0 | Finished |
| S68 | 6-0 | M | 0 | 0 | Finished |

Table VI

Errors of Fernald boys on 40 slides of the Intermediate Set (3U5)

(N-24)

| S | MA | CA | Errors First | Errors Repetitive | Outcome |
|---|---|---|---|---|---|
| S45 | 4-4 | 6-10 | 9 | 5 | Finished |
| S43 | 4-9 | 7-5 | 13 | 5 | Finished |
| S40 | 5-8 | 14-4 | 2 | 0 | Finished |
| S60 | 6-2 | 15-8 | 0 | 0 | Finished |
| S64 | 6-8 | 11-10 | 2 | 0 | Finished |
| S67 | 6-8 | 14-9 | 0 | 0 | Finished |
| S42 | 6-9 | 9-11 | 0 | 0 | Finished |
| S65 | 6-10 | 13-10 | 2 | 0 | Finished |
| S37 | 6-10 | 16-11 | 0 | 0 | Finished |
| S69 | 7-0 | 11-11 | 4 | 2 | Finished |
| S66 | 7-0 | 15-7 | 1 | 0 | Finished |
| S70 | 7-2 | 13-3 | 0 | 0 | Finished |
| S39 | 7-4 | 12-10 | 2 | 0 | Finished |
| S62 | 7-4 | 14-7 | 1 | 0 | Finished |
| S71 | 7-4 | 15-5 | 0 | 0 | Finished |
| S68 | 7-6 | 12-7 | 0 | 0 | Finished |
| S52 | 7-10 | 11-5 | 5 | 17 | Finished |
| S56 | 7-10 | 14-5 | 0 | 0 | Finished |
| S57 | 8-2 | 13-5 | 3 | 3 | Finished |
| S38 | 8-2 | 15-6 | 2 | 0 | Finished |
| S41 | 8-4 | 14-7 | 1 | 0 | Finished |
| S61 | 8-8 | 15-11 | 0 | 0 | Finished |
| S59 | 8-8 | 16-0 | 0 | 0 | Finished |
| S63 | 8-10 | 14-5 | 1 | 0 | Finished |

Turning now to performances on the Intermediate Set (3U5), Table VI shows the errors by 24 Fernald boys on these 40 slides. The 24 Ss shown in Table VI are included among the 27 who performed on the Elementary Set shown in Table I. The 9 Ss with mental age intelligence test scores between 4-4 and 6-10 are the same as the 9 Ss in Table I.

First-error frequencies were low for all Ss except S45 and S43. S45 was extremely active and restless, and on unit 1 of the Advanced Set he used his elbows to press on the windows. Both S45 and S43 completed the sequence.

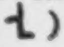
One of the older boys (S52) made 5 first errors and 17 repetitive errors. An analysis of the repetitive errors shows that 12 were made on slide 1-17 and all but one were made on the forms in windows 1 and 2. The sample was rotated 90 degrees to the right; one incorrect match (  ), in window 1, was also rotated 90 degrees to the

Table VII

Errors of Lesley-Ellis children on 40 slides of the Intermediate Set (3U5)

(N-9)

| S | CA | Sex | Errors<br>First | <br>Repetitive | Outcome |
|---|---|---|---|---|---|
| S54 | 4-0 | M | 1 | 0 | Finished |
| S55 | 4-3 | M | 1 | 1 | Finished |
| S64 | 4-5 | M | 2 | 0 | Finished |
| S70 | 4-6 | F | 0 | 0 | Finished |
| S52 | 4-10 | F | 2 | 0 | Finished |
| S60 | 5-8 | F | 0 | 0 | Finished |
| S61 | 5-9 | F | 0 | 0 | Finished |
| S67 | 6-0 | F | 2 | 0 | Finished |
| S68 | 6-0 | M | 0 | 0 | Finished |

right; and the correct match, in window 2, was rotated 30 degrees to the left. Responses to window 1 prior to window 2 occurred with high frequency.

Table VII shows the performances on the Intermediate Series (3U5) of the 9 Lesley-Ellis children who completed the Elementary Set shown in Table V. None of the children in this group made more than two first errors, and one S made one repetitive error.

Multi- and Unicolor Slide Series. Since the program was prepared in two forms, one multicolor ( all forms were of different colors within slides ) and one unicolor ( all forms were of the same color within slides but different between slides), information was obtained on the role of color variation within slides. The upper part of Table VIII shows the mean errors for 72 Fernald children, and the lower part, for 39 Lesley-Ellis children, on both series in the Elementary and Intermediate Sets. The over-all trends show that more errors are committed on multicolor slides.

### Performance of Fernald Children

Advanced series. Too few Lesley-Ellis children were observed on the last revisions of the Advanced Set to make meaningful comparisons with the Fernald children. Therefore, data on the Advanced series are given only on the performance of Fernald children.

Except for S43, all of the 24 boys in Table VI were continued on Advanced sequences. Thirteen were

Table VIII

Mean errors of Fernald and Lesley-Ellis children on unicolor and multicolor slides

|  | Unicolor Series |  |  |  |  | Multicolor Series |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|
|  | N | CA | MA | Errors<br>First | <br>Repetitive | N | CA | MA | Errors<br>First | <br>Repetitive |
| **Fernald** |  |  |  |  |  |  |  |  |  |  |
| Elem. (4U5)<br>24 slides | 28 | 6-10<br>to<br>16-11 | 4-4<br>to<br>8-10 | 1.5 | .4 | 10 | 10-7<br>to<br>15-1 | 5-3<br>to<br>8-4 | 2.4 | 1.9 |
| Inter. (3U5)<br>40 slides | 25 | 6-10<br>to<br>16-11 | 4-4<br>to<br>8-10 | 2.4 | 1.2 | 9 | 10-7<br>to<br>15-1 | 5-3<br>to<br>8-4 | 3.3 | 1.1 |
| **Lesley-Ellis** |  |  |  |  |  |  |  |  |  |  |
| Elem. (4U5)<br>24 slides | 9 | 4-0<br>to<br>6-0 |  | .5 | .0 | 11 | 3-5<br>to<br>5-10 |  | 3.0 | 3.2 |
| Inter. (3U5)<br>40 slides | 9 | 4-0<br>to<br>6-0 |  | .9 | .1 | 10 | 4-5<br>to<br>5-10 |  | 3.6 | .4 |

Table IX

Errors of Fernald boys on Unit 1, 46 slides, of the Advanced Set (5U5)

(N-10)

| S | MA | Age | Errors<br>First | <br>Repetitive | Outcome |
|---|---|---|---|---|---|
| S45 | 4-4 | 6-10 | 11 | 6 | Finished |
| S65 | 6-10 | 13-10 | 3 | 0 | Finished |
| S66 | 7-0 | 13-7 | 2 | 0 | Finished |
| S69 | 7-0 | 11-11 | 11 | 3 | Finished |
| S72 | 7-4 | 13-8 | 17 | 10 | Finished |
| S71 | 7-4 | 15-5 | 0 | 0 | Finished |
| S68 | 7-6 | 12-7 | 4 | 0 | Finished |
| S73 | 8-8 | 13-3 | 3 | 0 | Finished |
| S75 | 8-8 | 13-9 | 3 | 0 | Finished |
| S61 | 8-8 | 15-11 | 4 | 1 | Finished |

given unit 1 of the fourth revision of the series (4U5) in which there were five stages of change from non-mirror to mirror-images of the sample; and 10, unit 1 of the fifth revision (5U5), in which the transition was in six stages. Table IX shows that all 10 Ss finished the series, but 3 (S45, S69, and S72) showed a substantial number of errors.

The same 10 boys were given unit 2 of the Advanced Set, which contained forms with vertical plane rotations to 90 degrees. Their performances are summarized in Table X. S45 was stopped on slide 2-4 because of numerous errors and lack of progress. S72 was continued because his repetitive errors declined in approaching the end of the unit. The other Ss made fewer errors than S45 and S72 but more than they made on previous sets and units. On unit 3, which had vertical plane rotations to 135 degrees, 9 Ss were started; 4 were stopped and 5 finished. Their performances are shown in the upper part of Table XI. The five Ss who finished unit 3 also finished units 4 and 5, as shown in the middle and lower sections of Table XI. In addition to extreme vertical plane rotations, unit 4 had two, then

Table X

Errors of Fernald boys on Unit 2, 40 slides, of the Advanced Set (5U5)

(N-10)

| S | MA | Age | Errors<br>First | <br>Repetitive | Outcome |
|---|---|---|---|---|---|
| S45 | 4-4 | 6-10 | | | Stopped on slide 4 |
| S65 | 6-10 | 13-10 | 8 | 9 | Finished |
| S66 | 7-0 | 13-7 | 5 | 0 | Finished |
| S69 | 7-0 | 11-11 | 7 | 2 | Finished |
| S72 | 7-4 | 13-8 | 14 | 17 | Finished |
| S71 | 7-4 | 15-5 | 2 | 1 | Finished |
| S68 | 7-6 | 12-7 | 3 | 1 | Finished |
| S73 | 8-8 | 13-3 | 3 | 0 | Finished |
| S75 | 8-8 | 13-9 | 3 | 1 | Finished |
| S61 | 8-8 | 15-11 | 5 | 4 | Finished |

Table XI

Errors of Fernald boys on Units 3, 4, and 5, 40 slides each, of the Advanced Set (5U5)

| S | MA | Age | Errors<br>First | <br>Repetitive | Outcome |
|---|---|---|---|---|---|
| | | | Unit 3 (N-9) | | |
| S65 | 6-10 | 13-10 | | | Stopped |
| S66 | 7-0 | 13-7 | 13 | 5 | Finished |
| S69 | 7-0 | 11-11 | | | Stopped |
| S72 | 7-4 | 13-8 | | | Stopped |
| S71 | 7-4 | 15-5 | 11 | 8 | Finished |
| S68 | 7-6 | 12-7 | 4 | 0 | Finished |
| S73 | 8-8 | 13-3 | 3 | 0 | Finished |
| S75 | 8-8 | 13-9 | 11 | 5 | Finished |
| S61 | 8-8 | 15-11 | | | Stopped |
| | | | Unit 4 (N-5) | | |
| S66 | 7-0 | 13-7 | 5 | 1 | Finished |
| S71 | 7-4 | 15-5 | 8 | 2 | Finished |
| S68 | 7-6 | 12-7 | 0 | 0 | Finished |
| S73 | 8-8 | 13-3 | 4 | 0 | Finished |
| S75 | 8-8 | 13-9 | 13 | 6 | Finished |
| | | | Unit 5 (N-5) | | |
| S66 | 7-0 | 13-7 | 4 | 0 | Finished |
| S71 | 7-4 | 15-5 | 17 | 10 | Finished |
| S68 | 7-6 | 12-7 | 2 | 0 | Finished |
| S73 | 8-8 | 13-3 | 15 | 9 | Finished |
| S75 | 8-8 | 13-9 | 14 | 8 | Finished |

three, incorrect mirror-images of the sample, and unit 5 had four. In unit 5 all forms were the same in shape, all matches but the correct match were mirror-images of the sample, and forms were in one of 12 rotational positions.

As is apparent, several of the Fernald children made many repetitive errors. However, one boy (*S*25) was particularly persistent in repeating errors on unit 3 of the Advanced Set ( 5U5 ). In order to see what the course of his performance would be, he was allowed to continue to the end of the unit. Listed below are the number of repetitive errors in the order in which they were made. The numbers of slides in which there was one or zero errors are omitted.

| Slide Number | Repetitive Errors | Slide Number | Repetitive Errors |
|---|---|---|---|
| 4 | 1 | 39 | 9 |
| 4 | 2 | 38 | 23 |
| 5 | 3 | 37 | 2 |
| 4 | 2 | 39 | 1 |
| 7 | 1 | 39 | 5 |
| 7 | 4 | 39 | 2 |
| 18 | 22 | 38 | 2 |
| 18 | 7 | 37 | 1 |
| 21 | 1 | 38 | 5 |
| 20 | 1 | 38 | 1 |
| 22 | 5 | 37 | 8 |
| 26 | 3 | 37 | 53 |
| 30 | 1 | 36 | 4 |
| 35 | 1 | 38 | 1 |
| 34 | 1 | 37 | 16 |
| 33 | 1 | 38 | 2 |
| 35 | 1 | 37 | 6 |
| 37 | 2 | 37 | 1 |
| 40 | 1 | 40 | 1 |

In reviewing these figures, it is to be remembered that a repetitive error of one means that two errors were made on the same slide before the correct response was made, and a repetitive error of two means that three errors were made before the correct response, etc. Also, the reoccurrence of a slide number means that one or more errors were made on one or more previous slides. Finally, it is to be noted that one repetitive error was made on slide 40 the first time it was presented; after that, 143 errors were made on slides 37, 38, and 39 before slide 40 was responded to correctly on first presentation. He was exposed to the 40 slides in this series 116 times.

Mental Age Test Scores and Progress in the Program. The relationship between mental age intelligence test scores and number of slides completed in the program was evaluated on a group of 39 Fernald boys with mental age scores from 3 to 8. For this analysis, the criterion for stopping a *S* was three consecutive first errors. The mean number of slides completed for each mental age group is shown below:

| Mental Age Scores | *N* | Mean No. Slides Completed |
|---|---|---|
| 3 and 4 | 7 | 5.4 |
| 5 | 5 | 12.0 |
| 6 | 11 | 16.3 |
| 7 | 9 | 18.0 |
| 8 | 7 | 20.4 |

As is apparent, there is a positive relationship between test scores and progress through the program. The differences between successive mental age score means are not reliably different but the differences between the means for the three- and four-year group and the six -, seven-, and eight-year-old groups are statistically significant. The relevant figures are given below:

| Groups | *t* | d.f. | *p* for 2-tailed |
|---|---|---|---|
| 3-4 vs. 5 = | 2.70 | 10 | .05 |
| 3-4 vs. 6 = | 3.94 | 16 | .01 |
| 3-4 vs. 7 = | 3.98 | 14 | .01 |
| 3-4 vs. 8 = | 6.41 | 12 | .01 |

| Groups | t | d.f. | p for 2-tailed |
|---|---|---|---|
| 5 vs. 6 = | 3.44 | 14 | .01 |
| 5 vs. 7 = | 4.80 | 12 | .01 |
| 5 vs. 8 = | 2.69 | 10 | .05 |
| 6 vs. 7 = | .52 | 18 | .05<br>(not sign.) |
| 6 vs. 8 = | 1.36 | 16 | .05<br>(not sign.) |
| 7 vs. 8 = | .69 | 14 | .05<br>(not sign.) |

## DISCUSSION

The main findings from the construction of the form program (mainly described in the appendix) and from comparative analysis of Fernald and Lesley-Ellis children may be summarized as follows:

1. Effects of training on the form program. Both the Fernald and the Lesley-Ellis children made fewer errors on the Original, New Nonsense, and Alphabet tests after training. Hence, improvement was shown on material from the training program and on material that varied from it somewhat. There were no striking differences in the amount of improvement made by each group, and both groups performed about the same before and after training, making slightly more errors on the New Nonsense series.

2. Influence of the location of the sample with respect to the matches. During the pre-tests both groups made the highest number of errors on the matches displayed in the middle of five windows, the next highest on matches adjacent to the middle window, and the lowest on those in the end windows, giving a distribution over the five windows approximating a normal curve. During the post - tests, however, both groups showed a rather even distribution over all the windows. Subsequent research (Wolf & Bijou, 1963 ) showed that the location of the sample with respect to the matches was a significant condition influencing the selection of a match when the problem was extremely difficult. Errors could be increased on the first, middle, or last window by placing the sample over each of these windows.

3. Relationships between MA and CA and progress in the form program. There was a positive relationship between MA test scores and progress in the form program for the Fernald children. MA test score data were not available for the Lesley-Ellis children.

There was also a positive relationship between CA's and degree of advancement in the program for both the Fernald and the Lesley-Ellis children.

4. Properties of the programmed materials. Analysis of the performances of Lesley-Ellis children showed that of all the methods used to ease the children into discriminating between mirror-images and non-mirror-images of the sample, "the fading-in" technique was the most expedient. This approach is similar in principle to the one used by Terrance on infrahuman subjects ( 1963 ), and by Moore and Goldiamond on young children (1964).

Analysis of performances on the program of both groups showed that after discriminations were made to one mirror-image, new errors were caused by rotating and introducing additional mirror-images, which in many instances abolished the previously learned discriminations. In other words, once a mirror-image was discriminated from a non-mirror-image, there was no assurance that this discrimination would be maintained when the forms were rotated in the vertical plane and when additional mirror-images were displayed among the matches.

5. Influence of varying the color of the forms. In the series in which four colors were used in a slide, more errors were made than in the series in which only one color was employed. It seems that without special training, similarity of color in the sample and match

is a compelling condition for the selection of that match.

6. Effectiveness of the contingencies employed. No attempt was made to evaluate the effectiveness of the contingencies (light, chime, and progress in the program after a correct response, and "back-up" to the previous slide after an error) of a match-to-sample response or those involved in leaving the classroom or residential hall to come to the laboratory. However, data from casual observation suggested that more of the Lesley-Ellis children were concerned about making the correct matches, and more were reluctant to come to the laboratory for repeated sessions. On the other hand, it appeared that more of the Fernald children made more repetitive errors, displayed more extraneous behaviors ( e. g., nudging windows with the nose and fingering all windows before making a selection ), and expressed more enthusiasm for coming to the laboratory and more willingness to stay for longer periods. Subsequent work ( Heid, 1964 ) suggests that with some children, especially the young normals and the more deviant retarded ones, further progress in the program could have been achieved if more effective contingencies had been employed.

The implications of this research are several:

1. Programmed instruction techniques, now being developed by many investigators, can serve well as a training procedure to evaluate whether a skill or bit of knowledge not in the repertory of a child may be acquired. It is believed that the procedures described here can serve to develop programs for other types of concepts, including those described by Piaget in his sensory-motor, concrete thinking, and formal thinking operations ( Kessen & Kuhlman,1962; Piaget, 1954; Sigel, 1964; Smedslund, 1961) and by cognitive theorists relative to orientation and localization, perceptual constancies, form perception, and figure-ground relations (Wohlwill, 1960).

2. The approach used in this investigation has the promise of producing principles for the programming of nonverbal subject matter for educational purposes on at least three counts. It could lead to more effective learning of symbols, especially letters, words, phrases, and numbers. Skinner's exploratory work with letters and Thoft's (1963) investigation with geometric forms are cases in point (Holland, 1960). Second, it could provide impetus for sequencing of materials for the development of nonverbal skills not currently taught in schools. Reference here is to the study of processes such as " inductive reasoning " as attempted by Holland and Long (Holland, 1963 ), to analysis and synthesis of patterns of stimuli, and to figures and ground relationships and the like as developed by Long (1964). Third and finally, it could provide training in discrimination and abstraction that would facilitate the acquisition of constructed responses (e.g., writing, drawing, reciting ). Learning which helps an individual to discriminate the productions of others as well as his own provides him with a history that will help him reinforce his own behavior.

3. This approach also points to a way of obtaining information on the influence of historical variables on current performance(Bijou & Baer, 1963). Behavior of interest might be produced by experimentally constructing the necessary history. The end result of such an effort would be an unambiguous account of the variables and relationships of which such behavior is a function. An account of this sort would provide knowledge on one

possible way in which such behavior could be developed and maintained in the context of the usual circumstances of development. Jeffrey's work (1958a, 1958b) followed this strategy in evaluating the role of mediating behaviors in establishing a right-left discrimination and in differentiating tones separated by 3 octaves and a fifth. His subjects were young normal children.

4. Finally, this approach can have implications for the diagnosis and treatment of perceptual, symbolic, and language disabilities such as may be involved in reading retardation, aphasia, and articulatory difficulties (e.g., Holland & Matthews, 1963). The end phase of one or more programs could be administered to provide diagnostic information, and the complete programs could be given as treatment. Part of the procedure would, of course, be devoted to checking the effectiveness of possible reinforcers and the generalization properties of the stimulus materials. It is apparent that clinical procedures of this sort would not be designed to pinpoint the location and extent of neurological damage, but rather to discover the requirements for arranging and rearranging materials, procedures, and motivational variables to establish or recover lost functions (Bijou, 1964).

## APPENDIX

Stages in the development of the form program. In the initial composition of a slide (selecting the sample, matches, colors, and arranging the rotational positions) many factors were taken into account in order to increase the probability that the correct match would be made, and at the same time, have each slide contain an appreciable change in the direction of the criterion task. In lieu of well - established empirical principles, the construction of the first sets of slides was based on the following guide-lines.

1. Selection of sample, its rotational position and color.

A. The same sample was rarely presented on successive slides, and the succession of samples depended mostly on the phase of the training program. For example, when the first rotation of the sample was introduced, practically all seven forms were presented as samples tilted to the east or west 30 degrees from the vertical, or when variations in forms of mirror-image figures were introduced, the three forms capable of mirror-image confusions (٩, ʅ, and ⊥) served as samples an equal number of times in an irregular order.

B. Rotations of the sample from the upright position were introduced in gradual stages, and rotations to the east and west were introduced in a mixed order. It appeared that the rotational position of the sample introduced different levels of difficulty depending upon shape of the figure. For example, tilting of the cross (+)60 degrees to the east or west was less distracting than tilting the " winged " figure ( ⊥ ) to the same extent. These differences were taken into account in introducing a new rotational position of the sample in both mirror-image and non-mirror-image training.

C. Samples were not rotated to a new position until after the matches were presented in that position.

D. The color of the sample, from slide to slide, was varied in a random order.

2. Selection and location of the correct match.

A. The correct stimulus was never placed in the same window on successive occasions, on the assumption that the window containing the correct stimulus in the previous frame had, at least momentarily, greater stimulus control than those in the other four windows. Often, and depending on other factors, forms which differed markedly from the sample were presented in the window that was previously "hot."

B. It was assumed that the stimulus directly under the sample, in this case in the middle window, was most preferred, those on adjacent sides of the middle window were next preferred, and those in the end windows were least preferred. Window location, in terms of nearness to the sample, was exploited to increase the probability of a correct response to a new variation in stimulus dimension as well as to test the relative strength of nearness to sample over training in the matching on the basis of form and east-west orientation.

C. It was assumed that the location of a stimulus in the row of matches, independent of proximity to sample, also played a role. Figures similar to the sample in shape would be more likely to be responded to as correct if surveyed before rather than after encountering the correct stimulus.

The program and procedures did not train or force the *S*s to follow a fixed sequence. Rather than ignore the possible influence of location, however, matches were arranged on the assumption that most children moved from left to right. (Impressions from casual observation of eye movements suggested that six- and seven-year-olds were more likely to proceed from left to right than threes and fours.)

D. In the part of the program involving discrimination of the correct match from a mirror-image of it, the proximity of mirror-image and non-mirror-image matches was taken into account. The assumption was that discriminations are facilitated when similar forms are close to each other. ( See Figure 4, slides 1-35 and 2-2.)

3. Relationships among sample, correct matches, and incorrect matches.

A. It was assumed that the greater the rotational discrepancy between sample and correct match, the greater the difficulty of discrimination--at least within certain limits.

B. It was also assumed that the more similar the rotational deviation between the correct and incorrect matches, the more difficult the discrimination. In the part of the program which presented a non-mirror-image of the sample and two, three, and four mirror-images of it, each additional mirror-image was introduced in a parallel declination to the others and then gradually presented in nonparallel rotational positions. ( See Figure 4, slides 4-7, 5-3, and 5-38.)

C. Finally, it was assumed that the more similar the common elements ( location of the two squares relative to the vertical column of three squares ) between two figures, the more difficult the discrimination. This was held to be so whether the two figures involved a correct and an incorrect form or two incorrect forms.

*Pilot Studies.* Two preliminary studies were conducted to obtain information on workable size, color, and texture of stimuli, on techniques for preparing slides, on feasible units of rotational change,on experimental procedures, on time required for an *S* to complete units in the program, on amount of review necessary for a block of material, and the like.

The initial stimuli consisted of seven geometric forms similar to those in the left column of Figure 2. However, they were smaller (cut from 15/16-inch squares) and made of red, yellow. blue, and green blotter paper.

The first set of frames consisted of 95 slides with as sample always in the upright position, and four incorrect non -mirror-image matches made up of a combination of the following: + , [ , ⊥ , L , ₽ . The first frame had only one match. Additional matches gradually appeared on succeeding slides, and on slide 14 all five windows displayed matches. Rotational discrepancies between the upright position of the sample and matches varied gradually from 10 to 30 degrees. The forms were of different colors, but the colors of the sample and the correct match were the same only on those special occasions in which new rotational deviation was introduced. Color under these circumstances served as a prompt. Finally, there were two reviews, one in the middle of the program, with 12 slides, and one at the end, with 25.

Children from the Lesley - Ellis school served as Ss.

In the first pilot study there were seven children. Below are their ages and sex, the form of the program administered (combination of slides given ), the number of first errors, the number of repetitive errors, and the outcome.

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S5 | 3-3 | F | all 95 slides | No progress | - | Stopped |
| S6 | 4-0 | F | " | No progress | - | Stopped |
| S7 | 4-4 | F | " | No progress | - | Stopped |
| S4 | 5-9 | F | " | 2 | 0 | Finished |
| S2 | 5-11 | F | " | 1 | 0 | Finished |
| S1 | 6-4 | M | " | 5 | 5 | Finished |
| S3 | 6-5 | M | first 5 & last 14 | 1 | 0 | Finished |

The sequence was obviously too easy for Ss 5-9 and above, and too difficult for those 4-4 and below. S3 made the last 14 matches correctly without training on the previous 76 slides.

In the second pilot study, 50 slides were eliminated, including all review items. The revised program of 45 frames was administered to four Ss. Data from them suggested that the matches were still too easy for five-year-olds, that the review items were not necessary, and that the sample need not be maintained in the upright position.

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S8 | 4-11 | M | all 45 slides | 6 | 11 | Finished |
| S11 | 5-5 | M | first 14 & last 10 | 2 | 6 | Finished |
| S9 | 5-7 | M | all 45 | 3 | 3 | Finished |
| S10 | 5-8 | F | first 14 & last 10 | 0 | 0 | Finished |
| S12 | 5-8 | M | all 45 | 0 | 0 | Finished |

On the basis of data from the two pilot studies, the following changes were made:

1. A new set of geometric figures was prepared. These were larger (cut from 1 1/2-inch rather than 15/16-inch squares ) and were made of glossy rather than blotter paper. The glossy forms produced more vivid colors on the windows.

2. Three new sequences were prepared -- Elementary, Intermediate, and Advanced. The Elementary Set, which was for children as young as three, began with a ball, solid square, and triangle, since it was shown that these forms are discriminable to young babies of about six months of age ( Ling, 1941 ). The Elementary Set was also designed to strengthen the response required in matching to sample and to train in matching forms in the presence of the irrelevant cues produced by simple vertical plane rotations. The Intermediate Set was planned to provide training in matching on the basis of form despite vertical plane rotations to 180 degrees. The Advanced series

was designed to give further training in matching on the basis of form in the context of stimuli with mirror-images of the sample and differences in rotational positions to 180 degrees. All rotational changes were in units of 30 degrees.

3. The programs were prepared in two versions: in one, the forms were of different colors (Form M); in the other, they were of the same color (Form U). That is, in Form M, the colors were mixed within a slide; in Form U, colors were mixed between slides.

The Intermediate Set. The first version of the Intermediate Set contained 64 slides with seven forms (ʅ, +, [, ⊥, L, ɫ, P) and prepared in two series, a mixed-color (Form M) and a unicolor (Form U) sequence. The program began with a single match ( "forced choice"), but by slide 8 all five matches appeared. (The transition from one to five matches was made in 14 slides in the pilot studies.) Tilting of the sample and matches began in slide 5, and by slide 8, the sample and matches were at 30 and 60 degrees to the east (plus) and west (minus) of the vertical. From slides 9 to 21 the choices were 60 and 90 degrees, plus and minus, with four forms (ʅ, P, +, L) serving as samples in the upright position. From slides 22 to 41 the same four samples were gradually rotated through plus and minus 90 degrees, and from slides 42 to 64, the end of the set, rotations were extended to plus and minus 180 degrees and all seven forms appeared as samples. On the basis of the performances of four children, it was clear that the series was too easy for five- and six-year-olds.

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S15 | 5-6 | M | M | 3 | 0 | Finished |
| S16 | 5-7 | M | U | 1 | 2 | Finished |
| S13 | 6-7 | M | M | 2 | 0 | Finished |
| S14 | 6-7 | F | U | 0 | 0 | Finished |

Eighteen slides were eliminated and four other six-year-old children were tried, yielding the following results:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S27 | 6-3 | M | M | 13 | 11 | Finished |
| S28 | 6-7 | F | U | 0 | 0 | Finished |
| S30 | 6-10 | M | U | 2 | 0 | Finished |
| S29 | 6-11 | M | M | 4 | 0 | Finished |

Since the elimination of 18 frames increased the frequency of errors, they were restored and the Intermediate Set was used in its first version while the Elementary and Advanced Sets were being developed, despite the fact that it appeared to be too easy for the five- and six-year-olds.

Further evidence that the paring was too drastic for the three- and four-year-olds and not drastic enough for the five- and six-year-olds was supplied by data gathered while developing the Elementary and Advanced Sets. For example, three of the youngest children fared as follows:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S31 | 3-7 | M | U | 6 | 9 | Stopped |
| S32 | 3-10 | F | U | 3 | 4 | Finished |
| S33 | 4-5 | M | U | 3 | 2 | Stopped |

On the other hand, 15 five- and six-year-olds went through the sequence, after completing the Elementary, without difficulty, 12 without a single error and 3 with errors as follows:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S46 | 5-7 | F | U | 5 | 5 | Finished |
| S45 | 5-11 | F | U | 2 | 2 | Finished |
| S2 | 6-1 | F | U | 3 | 0 | Finished |

*The Elementary Set*. The original plan was to have the three- and four-year-olds begin with the Elementary Set and the five- and six-year-olds with the Intermediate. In the final version, however, all children were started with the Elementary, the upper end of which was made to dove-tail with the beginning of the Intermediate.

The Elementary Set, also prepared in mixed (Form M) and unicolor (Form U) forms, started with a ball 1 1/2 inches in diameter as sample and one match. The next three slides were the same as the first except that the single match appeared in different windows to the left (windows 1 and 2) and to the right (windows 4 and 5) of the middle window. The fifth and sixth frames also had single matches, but the solid square replaced the ball. These two slides were followed by a block of seven slides ( 7 to 13 ) with two matches; and the ball, square, and triangle as samples. In the next block of nine slides (14 to 22), three matches were available, and at the same time the experimental forms were gradually substituted for the ball, square, and triangle. By slide 18 the circle, square, and triangle were eliminated. Four matches appeared in slide 23, and five in slide 26. The matches were first rotated in slide 23, and by slide 29, the end of the program, all were in rotated positions.

The performances of three preschool children were as follows:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S18 | 3-5 | M | U | 3 | 25 | Stopped |
| S17 | 4-4 | F | U | 7 | 12 | Stopped |
| S19 | 4-6 | M | U | 5 | 18 | Finished |

These data, together with impressions gained from observation, suggested that the placement of single matches on either side of the sample in the first six slides strengthened alternating behavior which interfered with looking and responding to the form property of the stimuli.

The sequence was therefore revised in three ways. (1) The number of repetitions of forced matches was reduced. (2) The appearance of five matches occurred in slide 17 rather than in 29. (3) The total number of slides was reduced from 29 to 20.

Four preschool boys finished the program with fewer errors than before, as shown below:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S20 | 3-5 | M | 2U | 0 | 0 | Finished |
| S21 | 3-9 | M | 2U | 0 | 0 | Finished |
| S22 | 4-3 | M | 2U | 1 | 0 | Finished |
| S23 | 4-3 | M | 2U | 1 | 0 | Finished |

Based mostly on changes in behavior other than error frequency (*e.g.*, gradual shifts in topography and in latency), the sequence was further revised to allow the child to move ahead more rapidly. This was accomplished by starting with the ball as sample and three matches (ball, square, and triangle), and by presenting four matches in slide 4 rather than in slide 15, and five matches in slide 6 rather than in slide 17. Results on three children were as follows:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S31 | 3-7 | M | 3U | 1 | 0 | Finished |
| S32 | 3-10 | F | 3U | 1 | 0 | Finished |
| S33 | 4-5 | F | 3U | 3 | 3 | Finished |

This version of the Elementary Set (Elem. 3) was retained until the three sets were integrated.

*The Advanced Set*. The Advanced Set was designed to train the child to discriminate mirror-images of a sample as incorrect matches. Up to this point, incorrect matches were of different shapes with or without vertical plane rotations.

Only three of the seven geometric forms could serve as samples for mirror-image discrimination -- the P, ʃ, and ɟ.

In the first version of the program, the first slide had the as sample and four mirror-images of it as incorrect matches, all in the upright position. The next two slides also had the P as sample and the same combination of choices, only in different windows. This pattern was repeated for the ʃ and for the ɟ. In other words, each form served as sample in blocks of three. There was a total of 59 slides including 15 review items.

Performances of 6 youngsters between 4 and 6 years of age were as follows:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S22 | 4-3 | M | U | 3 | 2 | Requested to stop |
| S19 | 4-6 | M | U | 3 | 10 | Stopped |
| S15 | 5-6 | M | M | 5 | 5 | Stopped |
| S16 | 5-7 | M | U | 5 | 0 | Finished |
| S14 | 6-7 | M | U | 10 | 24 | Stopped |
| S13 | 6-7 | F | M | 7 | 9 | Stopped |

Except for S16, none of the children made progress beyond the tenth slide. Mirror-image and non-mirror-image matches were responded to almost at random.

The opening phase of the program was therefore revised. (1) A new first slide was prepared which had only one incorrect mirror-image match. The other three incorrect matches differed from the sample in shape. (2) The three forms that served as sample were presented in a mixed order rather than in blocks of three.

As indicated below, these revisions did not permit three six-year-olds to make progress through the program. All errors were made on the first 10 slides.

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S24 | 6-6 | M | M | 8 | 24 | Stopped |
| S25 | 6-8 | F | U | 6 | 12 | Stopped |
| S26 | 6-11 | M | U | 2 | 5 | Stopped |

Two more changes were made: (1) The initial slides were changed so that the incorrect mirror-image matches were presented in rotational positions differing from the sample and correct match, and at the same time the sample and correct match were in similar rotational positions. (2) To reduce the influence of distractors, the first three slides had only two matches, one like the sample and one a mirror-image of it. In slides 4 to 6 one incorrect non-mirror-image (forms that differed from the sample in shape) was included; in slides 7 and 8 two more were added; and in 9 three were included. Hence, from frame 9 on, the matches consisted of the correct match, one incorrect mirror-image of the sample, and three incorrect matches with forms differing from that of the sample.

As indicated by the performances of four Ss listed below, these changes still did not permit the children to make progress.

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S34 | 5-10 | F | 2 to 5 matches | 8 | 8 | Stopped |
| S35 | 6-1 | M | 2 to 5 matches | 4 | 0 | long latencies; stopped |
| S36 | 6-2 | M | 5 matches | 6 | 7 | Stopped |
| S37 | 6-5 | M | 5 matches | 2 | 2 | long latencies; stopped |

The fourth revision took a different tack. A new set of slides was prepared in which the three mirror-image forms were presented after exposure to forms that became mirror-images after construction in four stages. The mirror-image of _⌈ was gradually formed by lowering the right wing of ⊤ to the bottom of the figure, and the mirror-image of ⌉_ was developed by dropping the left wing of ⊤ . The mirror-image of _⊢ was evolved from dropping the right wing of the + to the base of the figure, and the mirror-image of ⊣_ by lowering the left wing of + . The mirror-image of P was evolved by raising the lower horizontal bar of ], and the mirror-image of ꟼ by raising the lower horizontal bar of [ . (These were the first approximations of those shown in Figure 4.)

Revision four contained other features. The three critical forms were displayed sometimes oriented east and sometimes west. Furthermore, window location of pre-mirror-image stimuli (those in construction stages) was taken into account. Figures easy to discriminate, such as those in the first and second construction stages, were originally presented with one non-mirror-image form separating them from the correct matches; figures difficult to discriminate, or those in the third and fourth construction stages, were initially presented in windows adjacent to correct matches.

One new variation of the entire program was prepared. It was like the unicolor series except that the matches were the correct form and the mirror-image of it. In other words, there were no matches which differed in shape from the sample. Notations in the table below indicate the form of the program as "U and 5" and "U and 2," which mean unicolor slides with five matches and unicolor slides with two matches, respectively. There were 32 slides in the five-choice version and 33 in the two.

Performances of four children were as follows:

| Subject | Age | Sex | Form | First Errors | Repetitive Errors | Outcome |
|---|---|---|---|---|---|---|
| S41 | 5-9 | M | U+5 | 0 | 0 | Finished |
| S39 | 5-11 | M | U+2 | 3 | 0 | Finished |
| S38 | 6-3 | F | U+2 | 1 | 0 | Finished |
| S40 | 6-4 | M | U+5 | 0 | 0 | Finished |

All four of the children completed the sequence, making a total of four first errors. Differences between performances on the two-choice and five-choice sequence were not apparent in this first run.

In order to see whether training on the pre-mirror-image forms facilitated performances on the earlier version of the program, the four children above were given the previous version of the program ( Adv. 3 ). Two of the four, S38 and S39, had the unicolor sequence with two choices ( U2 ), while S40 and S41 had the unicolor sequence with five choices (U5). All four made so many errors that they were stopped in the initial part of the program again.

The Advanced 4 version was gradually expanded. In the first extension the number of slides was increased from 32 to 51. Slides with smaller variations were placed in front of frames with many errors, and new slides were added with gradual increases in (1) rotational discrepancies and (2) window separations between correct matches and incorrect mirror-image choices. The second extension introduced two, three, then four incorrect mirror-images, and extended vertical plane rotations to 180 degrees. The total number of slides was increased from 51 to 166.

To summarize, the first 40 slides (unit 1) of the Advanced Set (4U5) had forms in four stages of mirror-image construction; the second 40 (unit 2) had one incorrect mirror-image among the matches and all forms were rotated more extensively than in unit 1; the third and fourth 40 (units 3 and 4) had two incorrect mirror-images; and the fifth 40 (unit 5) had three and four incorrect mirror-images with extremely wide-angle vertical plane rotations. High error frequencies on 4U5 were clustered in units 1 and 3. In 1 they centered around the slides with the first full mirror-images of the samples; and in 3, on rotational discrepancies at 90 degrees.

In the fourth revision (5U5) an attempt was made to reduce the erros on units 1 and 3. Unit 1 was extended to make more gradual the transition from incorrect non-mirror-images to incorrect mirror-images--full mirror-images evolved from five rather than four stages of change. Unit 3 was also extended by adding rotations at 75 degrees between the block of slides with 60- and 90-degree rotational discrepancies.

*Integration of the Elementary, Intermediate, and Advanced Sets.* Rather than starting three- and four-year-olds on the Elementary Set and five-year-olds and older on the Intermediate, the program and procedure were altered to start all *S*s on the Elementary Set. The Elementary Set was extended by four slides taken from the beginning of the Intermediate Set, and the Intermediate Set was reduced from 64 to 40 slides. The beginning of the Intermediate Set was made continuous with the upper end of the Elementary Set. The Advanced Set was made continuous with the upper end of the Intermediate Set. The last phase of the Advanced Set contained slides with four incorrect mirror-images in any of 12 rotational positions in the vertical plane. ( See Figure 4, slide 5-38. ) The integrated program consisted of 270 slides -- 25 in the Elementary Set, 40 in the Intermediate, and 205 in the Advanced.

## REFERENCES

Bijou, S. W. Research in behavior modification of normal and deviant children. In L. Krasner & L.P. Ullmann (Eds.), *Research in behavior modification: new developments and their clinical implications.* New York: Holt, Rinehart, & Winston, 1964.

Bijou, S. W. & Baer, D. M. Some methodological contributions from a natural science approach to research in child development. In L. P. Lipsitt & C. C. Spiker (Eds.), *Recent advances in child development and behavior*. Vol. I. New York: Academic Press, 1963.

Gesell, A. & Amatruda, Catherine S. *Developmental diagnosis: normal and abnormal child development.* New York: Hoeber, 1941.

Heid, W. H. Non-verbal conceptual behavior of young children with programmed material. Unpublished doctoral dissertation. Univer. of Wash., 1964.

Hively, W. Programming stimuli in matching to sample. *J. exp. Anal. Behav.*, 1962, *5*, 279-298.

Hively, W. Parametric experiments with matching to sample. Unpublished manuscript, 1964.

Holland, Audrey & Matthews, J. The use of a teaching machine in training speech sound discrimination for articulation therapy. *Asha*, 1963, *5*, 474-482.

Holland, J. G. Teaching machines: An application of principles from the laboratory. *J. exp. Anal. Behav.*, 1960, *3*, 275-287.

Holland, J. G. New directions in teaching-machine research. In J. Coulson (Ed.), *Proceedings of the conference on applications of digital computers to automated instruction.* New York: John Wiley, 1963.

Jeffrey, W. E. Variables in early discrimination learning: I. Motor responses in the training of a left - right discrimination. *Child Develpm.*, 1958a, *29*, 269-275.

Jeffrey, W. E. Variables in early discrimination learning: II. Mode of response and stimulus difference in the discrimination of tonal frequencies. *Child Develpm.*, 1958b, *29*, 531-538.

Kessen, W. & Kuhlman, Clementina (Eds.). Thought in the young child. *Monog. of the Soc. Res. Child Develpm.*, 1962, *27*, No. 2.

Ling, B. C. Form discrimination as a learning cue in infants. *Comp. Psychol. Monogr.*, 1941, *17*, No. 2.

Long, E. A. Unpublished report. Univer. of N. Carolina, 1964.

Moore, R. & Goldiamond, I. Errorless establishment of visual discrimination using fading procedures. *J. exp. Anal.* Behav., 1964, *7*, 269-272.

Piaget, J. *The construction of reality in the child.* New York: Basic Books, 1954.

Sigel, I. E. The attainment of concepts. In M. L. Hoffman & Lois W. Hoffman (Eds.), *Review of child development research*. Vol. I. New York: Russell Sage Found., 1964.

Smedslund, J. The acquisition of conservation of substance and weight in children: VI. Practice on continuous versus discontinuous material in conflict situations without external reinforcement. *Scand. J. Psychol.*, 1961, *2*, 203-210.

Tanner, J. M. & Inhelder, Barbel (Eds.). *Discussions on child development*. New York: Internat. Univer. Press, 1960.

Terrace, H. S. Discrimination learning with and without "errors." *J. exp. Anal. Behav.*, 1963, *6*, 1-27.

Thoft, Judith S. Facilitation of form discrimination by experimental manipulation of an extraneous property. Unpublished master's thesis, Univer. of Wash., 1963.

Wohlwill, J. F. Developmental studies of perception. *Psychol. Bull.*, 1960, *57*, 249-288.

Wolf, M. M. & Bijou, S. W. Influence of the location of the sample in a match-to-sample problem. Unpublished manuscript, Univer. of Wash., 1963.

# FOOTNOTES

1. This research was accomplished while the investigator was on a Special Senior Fellowship from the National Institutes of Health (1961-62) under B. F. Skinner and James G. Holland, Harvard University. Subsequent work was supported by a grant (M-2232) from the National Institute of Mental Health, Public Health Service.

2. The courtesies extended by Miss Wright, principal of the Lesley-Ellis School, and Dr. Malcolm Farrell, superintendent, and Dr. John Ogonik of the Fernald School are gratefully appreciated.

3. The services of Mrs. Harriet Schultz and Mrs. Judith S. Thoft, who served as experimenter at the Lesley-Ellis School and the Fernald School respectively, were invaluable in exploring new procedures and in bringing the project to a conclusion.

4. Commercially referred to as "Color-aid" paper, produced by the Color-aid Company, 329 East 29 St., New York 16, New York. The yellow was designated #9, the red #57, the blue #130, and the green #170.

PREC.DING PAGE BLANK-NOT FILMED

# AN INVESTIGATION OF KEY-TONE MATCHING WITH CHILDREN AND ADULTS

### Edward Maltzman

The differentiation of one musical tone as "higher" or "lower" than (or, "same as" or "different from") another musical tone is typically called a discrimination. In the same way, the identification of a given tone by its musical staff letter name, or pitch frequency, without the aid of any previously given reference tone, is a discrimination. These two forms of discrimination are commonly identified as "pitch discrimination" and "perfect (or "absolute") pitch" (Lundin, 1963). Between these two points lies a discrimination area we may call "relative pitch." Discriminations in the relative pitch area include: the identification of the "distance" between two tones, e.g., a third, or semitone; and the identification of one tone, say C, from the standpoint of a previously given reference tone, say G. "Pitch discrimination," "perfect pitch," and "relative pitch" constitute generalized definitions of three major points or levels along the pitch discrimination continuum. In common usage, these broad definitions apply to levels of pitch discrimination "ability" or to special qualities of "hearing." ("Ability" is used here as a generic term to include any of the inner construct viewpoints of "musical ability," such as Seashore's [1919, 1938] "inner capacities" and Mursell's [1937] "mental functions." Implied in these definitions are the notions that "pitch discrimination" involves a single dimension along a single continuum, and that pitch discrimination ability is the operation of some inner state or function. However, since in any pitch discrimination situation, tones are involved (which may be close or far apart), and specific procedures (e.g., tones may be presented successively or simultaneously), as well as overt or covert behaviors (such as singing, speaking, "playing," "thinking"), it is evident that pitch discrimination involves, in addition to tones, a continuum of procedures, a continuum of behaviors.

Clearly, research into the phenomenon of musical learning has been in need of a new theoretical framework, new approaches, new methods of investigation, new data. A new operational framework for studying musical behavior became available with the development of the educational technology of programmed instruction initiated by B.F. Skinner (Skinner, 1954). The salient features of this new technology have been presented in a number of articles by Skinner and his associates (Skinner, 1954, 1958, 1961; Holland, 1960), as well as by a number of other authors (Lumsdaine and Glaser, 1960).

The critical features on which programmed instruction is based may be summarized as follows: (a) specification of the terminal be-

---

This article is based on a thesis presented to the Committee on the Ed.D. in Education, Boston University, January, 1964.

havior, (b) arranging for the student to emit behavior under controlled conditions, (c) arranging the material or skills to be learned in a gradual progression of learning steps, (d) providing immediate reinforcement for correct answers, (e) modifying the program to fit the student,and (f) permitting the student to progress at his own rate. Although programmed instruction frequently employs the use of a "teaching machine," such devices are not the defining characteristic of this technology. Teaching machines are not simply "for the mechanization of teaching, but instead the development of a new technology--a behavioral engineering of teaching procedures"(Holland, 1960, p. 275). Teaching machines are essential in providing for the precise control of contingencies the technology requires.

The use of this technology is not limited to application on the instructional side of learning. Equally significant is its use as a research tool for studying human learning (Hively, 1962; Porter, 1960). It is mainly in this latter application that the technology was employed in this study.

Although it has received only limited application in research practice,the notion of behavior as the criterion for learning music is not new.For example, Mursell noted that "the basic datum with which we have to deal is the phenomenon of behavior change" (Mursell, 1958, p. 140).However, with the traditional emphasis on the training of "ability," subtle relationships that occurred between behavior and the consequences of behavior were frequently overlooked, and significant variables were undefined and, hence, uncontrolled. But "ability" or "auditory imagery" cannot be observed or directly dealt with. Under the new technology, such terms "are defined by reference to the behavior from which we infer their presence" (Skinner, 1961, p.383).

This study, then, brought to the investigation of pitch discrimination certain objective methods of behavioral psychology. It attempted to deal with behavior directly, and was conducted with reference to the effects of certain tasks and procedures on the performance of individual subjects. The conditions which set the stage for behavior were largely controlled and recorded by means of a specially designed apparatus. Experimental variables were similarly controlled in an automated manner.

Purpose of the study. The broader purposes of the study were (a) to develop a methodology by which behavioral principles could be investigated in the area of pitch discrimination learning, and (b) to explore the feasibility and implication of a technology of teaching music based on behavioral principles. The specific purposes were:

1. To devise a key-to-tone matching task initially difficult for preschool and elementary school children.
2. To alter conditions within the procedure in the attempt to teach the behavior directly and objectively.
3. To study the effects of the procedure developed with youth subjects on musically naive adults.
4. To obtain objective accounts of some aspects of musical behavior.

The behavior selected for investigation was key-to-tone matching, i.e., "playing" appropriate organ keys to reproduce, thereby "matching," their correlated tones which had just been sounded. The task employed three key-tone associations in the familiar "do, re, mi" configuration. The procedure presented the individual tonal stimuli in successive matching trials. Experimental variables manipulated were (a) progression (sequence) of material, and (b) reinforcing stimuli.

Questions to which the study was directed included:

1. Would the "do, re, mi" configuration in successive matching trials constitute a "difficult" or "easy" musical task for individuals (as indicated by error rate)?

2. Could key-to-tone matching behavior be improved without having to employ other procedures (or other response modes)?

3. What behavioral characteristics would objective data reveal with respect to this musical task?

Organization of the study. The study was divided into an exploratory phase and a control experiment. The exploratory phase was concerned with testing tasks and procedures and with obtaining objective accounts of behavior. This phase consisted of: (a) Study 1A, which tested the general procedure of key-to-tone matching, with the "do, re, mi" stimuli in "random" successive matching trials; (b) Study 1B, which tested the effects of an "ordered" arrangement of the stimuli; (c) Study 1C, which tested a second randomized presentation of the stimuli as a check on error rate data yielded above; (d) Study II, which tested the discriminative values of a number of large to small interval sizes; and (e) Study III, which attempted an instructional program to teach the "do, re, mi" key-to-tone matching task.

The control experiment was based on the instructional program developed in Study III and studied the effects of three feedback and reinforcement procedures.

Errors as functions of trials, and operational relationships between stimuli and response events, constituted the data collected. These data were plotted in the forms of cumulated error curves and stimuli-response tracking flowcharts, to provide objective accounts of individual performances.

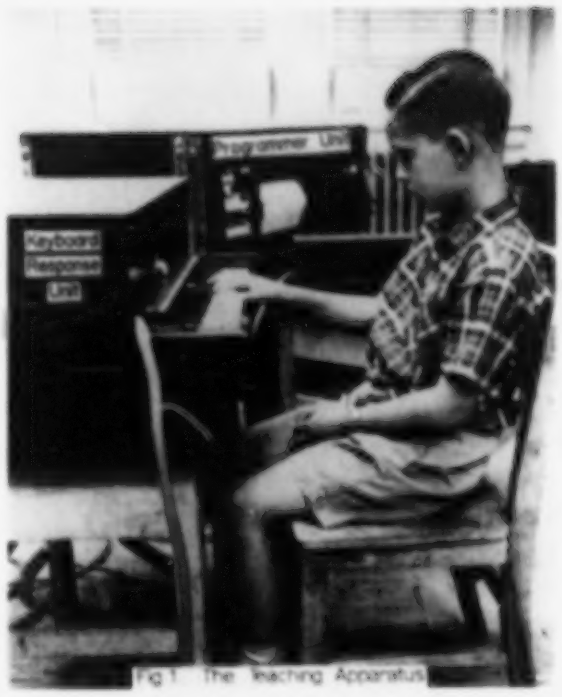


Fig 1 The Teaching Apparatus

An Implications and Suggestions for Further Research section concludes the study.

## EXPLORATORY STUDIES

Apparatus. The apparatus developed for this study is shown in Figure 1. It consisted of two integrated units. A three-octave keyboard and metal reed organ tone source were adapted to function as the Keyboard Response Unit. A piano player mechanism was adapted to function as the Programmer Unit This unit controlled stimuli presentations and response contingencies on the keyboard by means of a punched paper tape and appropriate relay circuitry.One set of punches stopped the tape for the presentation of a stimulus item. Another set of punches set up the response contingency in the keyboard, simultaneously setting up prearranged contingencies for feedback and reinforcement. A device wired to the keyboard recorded all key press responses. The apparatus automatically presented stimuli, recorded responses, and provided immediate feedback.

Subjects. Subjects were preschool and elementary school children from the Lesley Ellis School in Cambridge, Massachusetts, a private school attended by children from the academic and professional population in the Cambridge vicinity. Ages of subjects ranged from 4.2 to 11,5 years. No prescribed procedure was followed in the choice of subjects, who were initially selected by classroom teachers at the time of the experimenter's requests. Music training in the kindergarten consisted of singing and rhythmic activities handled by classroom teachers. In the elementary school, music was handled by a specialist; music consisted of singing, recorder playing, and listening activities for one twenty-minute period per week.

Study 1A. Study 1A utilized key-tone associations in the "do, re, mi" configuration in a successive key-to-tone matching procedure as a basic "musical" task. A response to the associated key in each trial produced the tone presented in that trial, simultaneously illuminating a red lamp signal mounted over the keyboard. A correct response also caused the tape to advance and the apparatus to present the next tonal item. A response to a key not associated with the presented tone constituted an incorrect response, which activated a buzzer sound. On an incorrect response, the tape did not advance and the subject could response again immediately.

Subjects (N = 8) were brought to the music room one at a time to "play a special game." When subject was seated at the keyboard, he was given the following instructions:

1. "This is a keyboard for you to play with. We will play with the keys I will now mark with crosses." E marked crosses on the middle octave keys--C, D, E--and started the machine.

2. "This is the way we play."

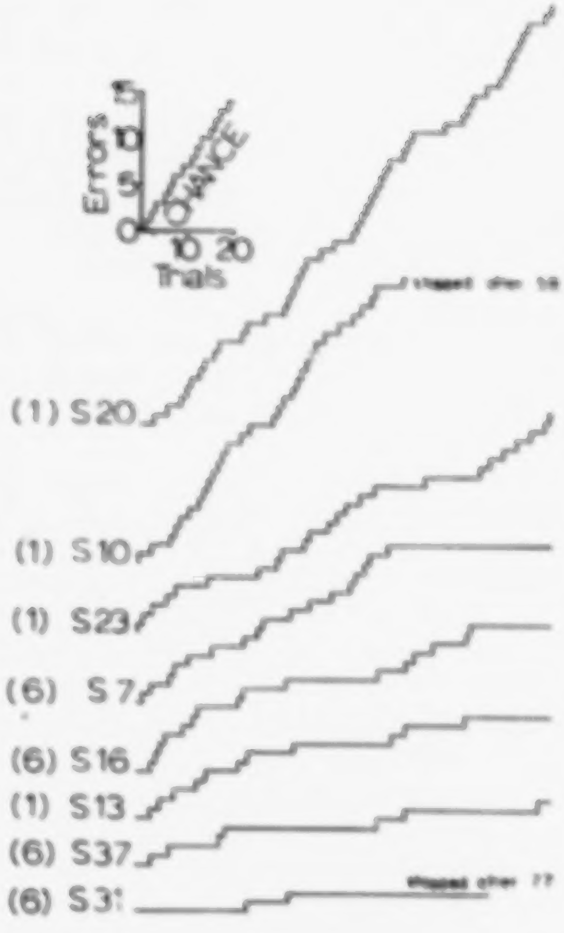


Fig 2. Cumulative records of 8 subjects on Series 3RS-90 items. Numbers in parentheses indicate grades of subjects.

E operated the machine correctly three times (making correct responses to items 1,2,3).

3. E reset the tape to item 1 and said, "You do it."

4. After subject responded to items 1, 2, 3, E said, "Good, see if you can play the right tone every time." E then took up a prearranged place behind the programmer unit, out of the subject's line of vision.

Results and discussion. Figure 2 presents cumulated error records for each subject on this series -- S10, S13, S20, S23, Grade 1; S7, S16, S31, S37, Grade 6. In these records, first response errors made by each subject were plotted cumulatively on the vertical axis against trials on the horizontal axis. The number of errors up to any point in a curve may be obtained by counting the number of vertical lines to that point. The responding is a horizontal line. The slope for chance performance is hown in Inset.

Figure 2 shows performances ranging from near perfect (S31) to chance (S20), with Grade 6 subjects making the better performances. S31 was stopped at item 77, since he obviously could do the task. S10's frequent production of the buzzer sound apparently became too aversive and she stopped working at item 58.

Study 1A served as a point of entry into this investigation of key-to-tone matching. Since past studies provided few clues with respect to characteristics of pitch discrimination behavior or to appropriateness of discrimination tasks for young children, Series 3RS was designed as a criterion task simply for the purpose of collecting data of individual performances which might provide an objective look into this behavior. This task and procedure differed from the types of tasks and procedures usually employed in studies involving pitch discrimination: instructions were minimal, related mainly to the operation of equipment and to the general procedure; behavior was limited to one mode of responding, namely, key pressing.

With the "do, re, mi" configuration so basic in Western musical culture, and the "do, re, mi" discrimination so generally assumed in instructional procedures, the number of errors produced was surprising.

In the better than chance performances, errors in the early part of the series may have been due to the process of adapting to the apparatus and to the procedure, that is, in "catching on" to the task; in the latter parts of the program, scattered errors might be attributed to such factors as "inattention." However, errors by sixth graders may also have indicated that these subjects were not secure in this discrimination. The performances of S37, S13, and S16 could be so interpreted. S7's record is indicative of learning. After producing errors through approximately two thirds of the series, she performed in the final portion with no errors.

The poorer than chance performances of S23, S10, and S20 require another interpretation. Were these children "pitch deficient" (Wolner and Pyle, 1933)? Or were these performances due to other variables operating in the procedure? Cumulative records do not provide sufficient information upon which to form such judgments. On the basis of casual observation, the behavior of these subjects appeared to be entirely random. For a closer look at an apparent chance performance, S20's data were plotted in flow chart form in which the sequence of stimuli presented and responses made in each item could be tracked simultaneously. A brief examination of this flow chart, shown in Figure 3, is sufficient to show that S20's errors were not randomly distributed but rather the result of systematic performance. Correct first responses were made in items 1, 2, and 3. In item 4, the first and second responses were incorrect; the third response was correct. Note, however, that the first incorrect response in item 4 was to the key that was the correct response in item 1, the initiation point of the "do, re, mi" sequence. Note, also, that the production of the correct response in item 5 may also have reinforced the 1, 2, 3 sequence of the first three items. That is, S20's second and third responses in item 4 and his first response in item 5 followed the sequence of responses in items 1, 2, and 3. In item 6, the subject's

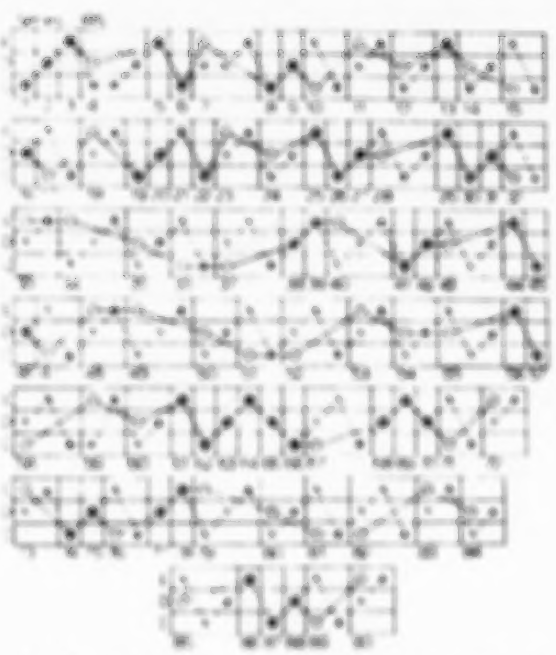


Fig. 3 Graph showing stimuli presented and responses made by S20 on Series 3RS (3 tones, random sequence)

Solid line connecting circles indicates sequence of stimuli

● Dot enclosed by circle indicates correct first response to stimulus

Broken line to x indicates incorrect response

Broken line to triangle indicates corrected response

first response again was to the key which was the correct response in item 4. Items 8 and 9 also apparently reinforced the 1, 2, 3 sequence pattern of responding which the subject was to follow throughout the balance of the series. This stimulus-behavior flow chart leaves no doubt as to what S20 did throughout the series. An interpretation of this performance must take into account the operation of sequence variables powerful enough to have kept the subject from coming under stimulus control. Subject's responses apparently were not based on what he was "hearing," but on the 1, 2, 3 sequential pattern. If such was the case, it would seem premature to attribute his performance to "pitch deficiency." Items 19-22 illustrate the occurrence of a number of correct responses for reasons other than stimulus control, i.e., when pattern of stimuli and pattern of responding happened to coincide.

Systematic behavior, expressed in somewhat different patterns, was similarly observed in flow charts of other subjects. The effect of systematic performance in the "presolution" period of learning was first noted by Krechevsky (1932) in visual discrimination studies with animals. Hively (1962) found this effect in visual matching-to-sample studies with children. To this writer's knowledge, this effect had not been previously highlighted in pitch discrimination studies.

These data provide a unique objective view of key-to-tone matching. In subsequent studies, certain of these data served as baselines against which changes in behavior were observed when changes were introduced into the procedure. A change in sequence of material was introduced in Study 1B.

Study 1B. In Study 1B the sequence of stimuli presentation was altered. It presented series 30S, a 35-item "ordered" sequence of the same tonal stimuli, followed by series 3RS. Six subjects were used in this study, four from the previous study (S7, S16, S20, and S23), plus two new subjects (S14 and S24) from the kindergarten. In all other respects the procedures were the same as in Study 1A.

Some questions which occurred were: Would the ordered series 30S produce fewer errors than the 3RS series? What effect would 30S have on a second performance on 3RS, particularly in the case of S20? For the two new kindergarten subjects, what effect would 30S have on their performance on 3RS?

Results and Discussion. Figure 4 presents cumulated error records of the six subjects in this study. S23 was given only the new series 30S. He produced the same number of errors in the new series as he did in the first thirty-five items in the previous series. This subject was not given a second try on

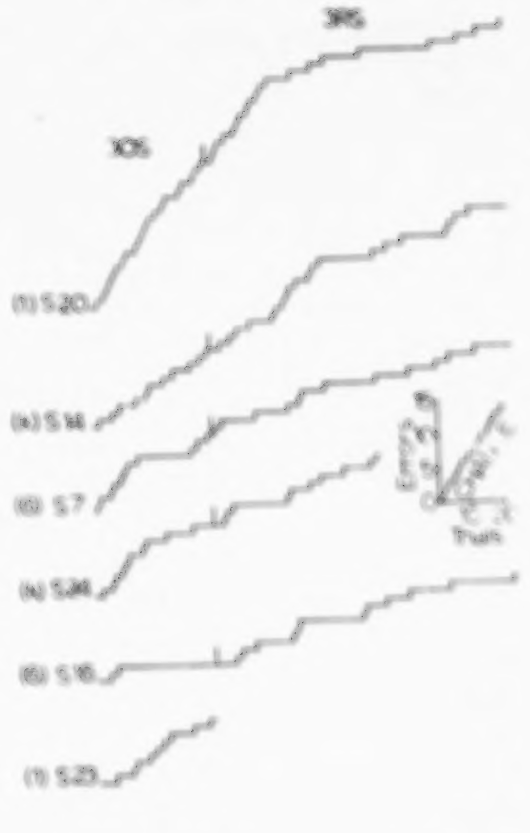


Fig 4. Cumulative records of 6 subjects on Series 30S - 35 items and 3RS - 90 items
(k) - kindergarten

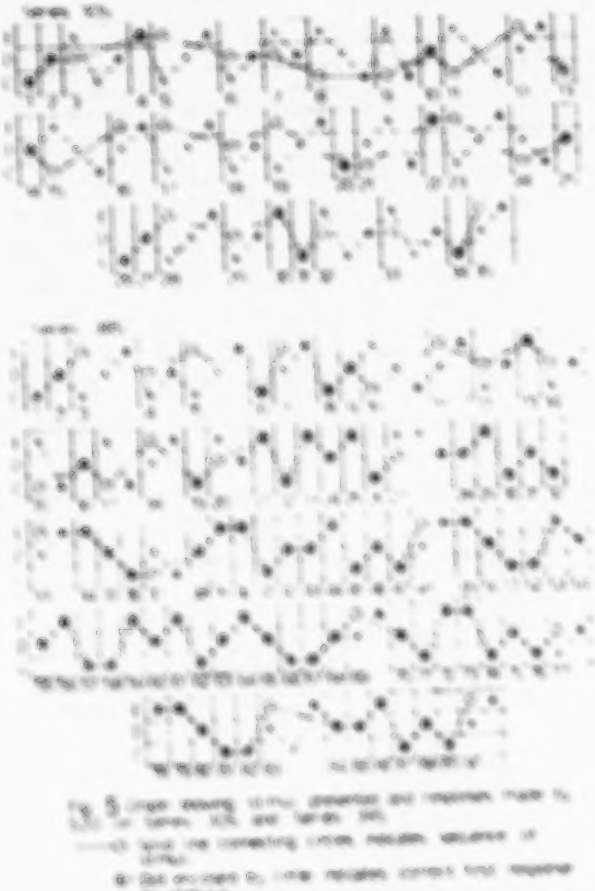


Fig. 5 Graph showing stimuli presented and responses made by S20 on Series 30S and Series 3RS.
[illegible]

3RS. From the number of errors produced on 30S, it would seem that this series was not very much "easier" than the random series.

S7, who in Study 1A produced a performance of learning, produced a similar performance in the new series (except for the cluster of three errors at the very end). But if she had learned the 3RS task in Study 1A, why was there no perfect performance the second time through? Series 30S apparently was "easier" for S16, but its effect on his second try on 3RS was negligible. S24 stopped working after fifty items in 3RS.

Although the effects of the new series appeared negligible, that behavior was undergoing changes was evident in S20's cumulated error records and stimulus-behavior flow charts. Examining his cumulated error records first (Figure 4), we see that his performance on 30S was at chance. In 3RS, his performance was at chance for the first twenty items, after which his curve flattened out considerably, indicating learning. His flow charts for these performances, shown in Figure 5, show that changes in behavior did occur. In 30S, we note that his previously extablished behavior pattern persisted in items 1 through 6. In items 7, 8, and 9, we see a breakup of the 1, 2, 3 alternation. The behavior pattern continued to fluctuate through items 20 in the following series 3RS, when stimulus control began to take over.

What the implications are of such an objective account of behavior have yet to be explored. However, these data indicate the danger of forming judgments as to any genetic "ability" or "auditory imagery" or "pitch deficiency" of a subject on the basis of an arbitrary test, since orderliness of behavior is not recorded in typical testing activities. That is to say, poor test results (poor performance) automatically attributed to variables inherent in, or otherwise connected with, the subject (such as "inattention" or "pitch deficiency"), might be shown by appropriate objective data to be systematic behavior, a function of unsuspected procedural variables.

Study 1C. Specifically, Study 1C was designed as a check on previous data, with respect to (a) the difficulty of the 3RS task, and (b) the apparent dominant influence of alternation with these stimuli.

Six new subjects were used: S73 and S71 were kindergarten children; S43, S69, and S61 were first graders; S72 was a sixth grader.

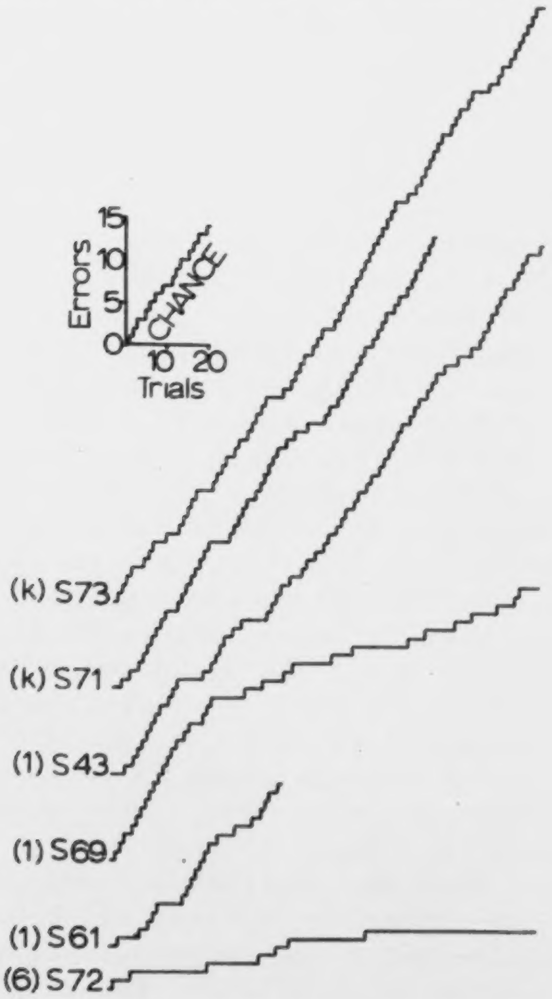


Fig 6. Cumulative records of 6 subjects on Series 3R2S - 100 items.

Results and Discussion. Figure 6 presents the cumulative records of these subjects on series 3R2S. Only one subject, S72 (sixth grader), had little difficulty with the task from the outset. The kindergarten and first grade children all did have difficulty. Only one record, that of S69, showed evidence of learning. S61 stopped at item 40;S71 stopped at item 75. A stimulus-behavior flow chart made of the performance of S73 on this series is shown in Figure 7. The influence of sequence of material presentation is immediately evident. The stimulus pattern in the first three items was **E**, **C**, **E**, to which correct first responses were made. In item 4,the first response was to the key that continued the **C-E** pattern. Notice the **E-C** preference occurring throughout the performance. Notice, too, the appearance of alternation in items 10, 11, 12. It is interesting that only in items 98 and 99 in this

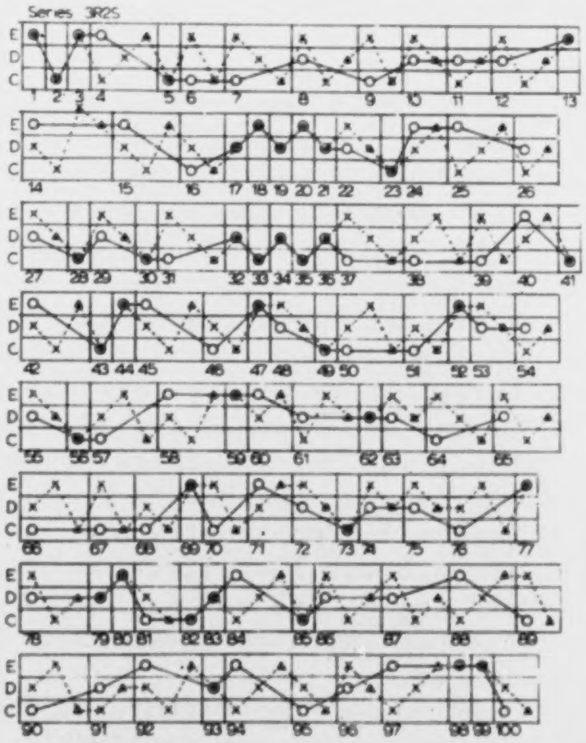


Fig 7 Graph showing stimuli presented and responses made by S43 on Series 3R2S

—○=Solid line connecting circles indicates sequence of stimuli
●=Dot enclosed by circle indicates correct first response to stimulus.
····x=Broken line to x indicates incorrect response
····▲=Broken line to triangle indicates corrected response

series did S43 correctly respond (in first correct responses) to stimulus repeats.

These results tend to support the previous data that (a)the "do, re, mi" discrimination in successive matching trials is not as "easy" for younger children as might be supposed, and (b) perhaps due to the natural 1, 2, 3 pattern of the stimuli themselves, or to the dominant characteristics of the procedure, a major form of irrelevant behavior appeared to be responding to the keys in the 1, 2, 3 order (alternation).

That this task appeared to be difficult for younger children does not necessarily mean that it constituted an inappropriate task for them. What of the stimuli themselves? Were they too close to be easily discriminated? In the following study the discriminative values of several large and small intervals in matching contexts with two tones were investigated.

Study II. Study II presented series 8TIS (eight tonal intervals sequence), shown in Figure 8, This series consisted of the following progression of interval sizes: eleventh, seventh, fifth, fourth, major third, minor third, major second, minor second, in which the same 25-item presentation sequence of interval tones was followed in each interval segment. The procedure described below for the eleventh segment was the same for all interval segments.

Since the material in this program changed from one interval size to another, a system of identification lights for the keys was installed over the keyboard to facilitate cueing procedures: ivory lights over the white keys, red lights over the black keys. (Cue light bar is shown in Figure 1.) These lights were manually controlled by E from a panel placed near the programmer unit. Key light cues are indicated in Figure 8 by letters enclosed in boxes over the items where cues were given.

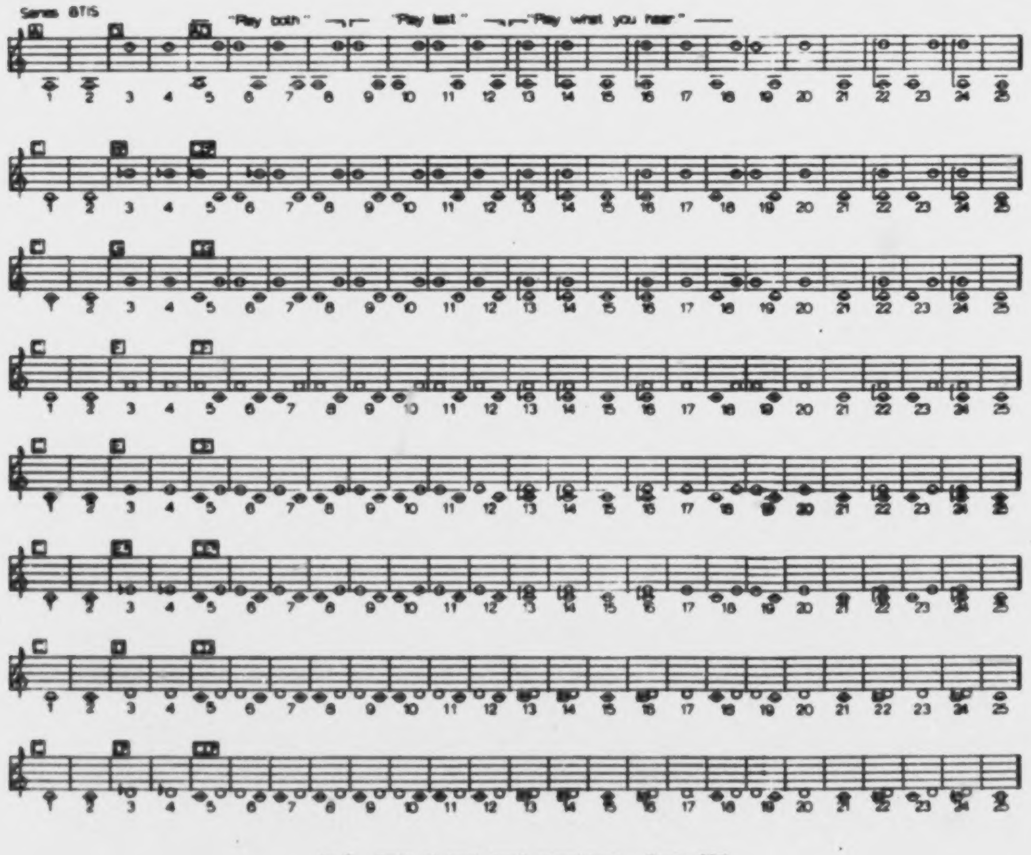


Fig 8 Order of tonal stimuli presentations - Series 8TIS

Pages 146, 147 and 148 have been omitted.

In item 1(in the top, the eleventh segment), the cue for key A was illuminated simultaneously with the presentation of tone A. The cue remained on for item 2, which presented the same tone. In item 3, the A key cue was extinguished and the cue for the D key was turned on simultaneously with the presentation of tone D. Items 1-4 required responding only to the lighted keys in each trial. In item 5, the D cue remained on and the A cue light added, both cues continuing through the balance of the interval sequence.

Items 5 through 8 presented the two tones, A and D, in melodic sequences; the subject's task was to respond on the A and D keys in the same sequence in which the tones were presented. No directions for playing the two keys in these trials were given unless a subject hesitated, in which case he was told, "Play what you hear." Only two subjects required this direction. At item 9, the subject was instructed, "Now play the second tone you hear." This instruction held through item 12. At item 13, the subject was instructed, "Play what you hear." Note that in items 13 and 14 both tones were presented simultaneously. Since these keys were widely separated, the use of both hands was required. (E repeated the instruction "Play what you hear" in only one case when the subject failed to respond immediately.) Items 15 through 25 presented the interval in the modes designated, calling for responses to the key or keys which would reproduce the tone or tones presented in each item. Feedback and reinforcement procedures were the same as in the previous studies; correct choices reproduced tones and illuminated red lamp signals; incorrect responses produced the buzzer sound. This procedure for the eleventh interval was followed in each of the succeeding interval segments. Subjects for this study included S10 and S20, who participated in earlier studies, and the following new subjects: S15, S36 from the first grade, S2, S25 from the second grade, and S35 from the third grade.

Results and Discussion. Figure 9 summarizes the performances of the seven subjects on this series. Errors occurring in the eleventh interval segment may be attributed to adaptation to the changing instructions. This figure indicates that for S36, S10, S15, and S2, discrimination became increasingly difficult as the interval sizes became smaller. S10 and S15 did not complete the series. S35 had some difficulty in the minor third segment, but recovered to produce a decreasing error rate in the balance of the series. S25 had no trouble throughout the series. The major and minor third segments were omitted in the case of S20 in view of his previous participation; ob-

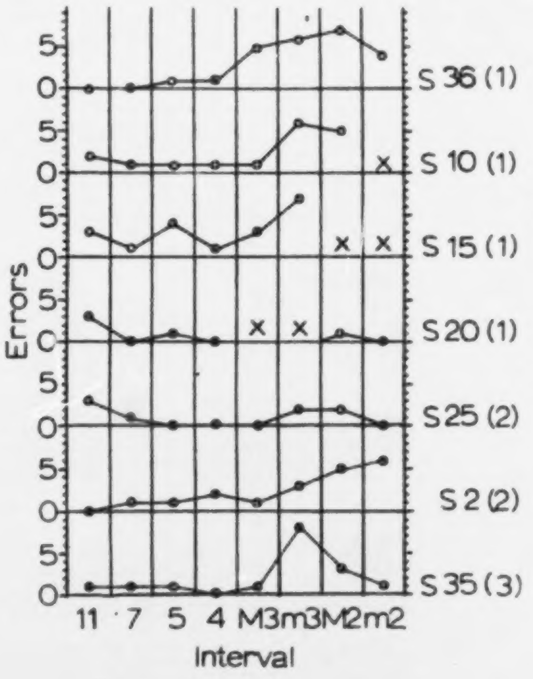


Fig. 9. Performances of 7 subjects on Series 8TIS.

x = Interval section omitted.

viously, he had no difficulty.

If the tones in this study are considered simply as stimuli along a tonal continuum, then these results are consistent with results of laboratory discrimination studies that stimuli at far points are more easily discriminated than stimuli at close points (Hilgard and Marquis, 1961, Chap. 12). On the other hand, if the stimuli are considered from the standpoint of their intervallic relationships, then these results are in apparent contradiction to other "musical" findings that whole tones and thirds are among the more easily discriminative intervals (Ortmann, 1934).

Although this study did not replicate any previous experiment, an analysis of errors suggests that the present results in fact are consistent with both types of findings. The increase in errors in the smaller interval segments supports the former finding. That errors did not occur in the trials in which intervals were presented, but did occur in the smaller interval segments in the single tone trials, is not inconsistent with the latter finding.

While knowledge of tonal discrimination behavior is admittedly limited, these findings suggest that this behavior is not only a function of the proximity of tones, but also of the procedure, and stimuli units employed. That is, a single tone represents a different stimulus unit than a melodic interval. It is likely that the discrimination behavior involved may differ in some essential way, depending on procedure and stimulus unit.

In the following study, an instructional program was attempted to teach the "do, re, mi" key-tone matching in single tone-matching trials.

Study III. Study III attempted an instructional program to train the series 3RS key-to-tone matching task. The contingencies for feedback and reinforcement were the same as in the previous studies. The program, shown in Figure 10, included a 25-item segment from series 3RS (transposed to F, G, A) as pretest and posttest, so that a measure of the effectiveness

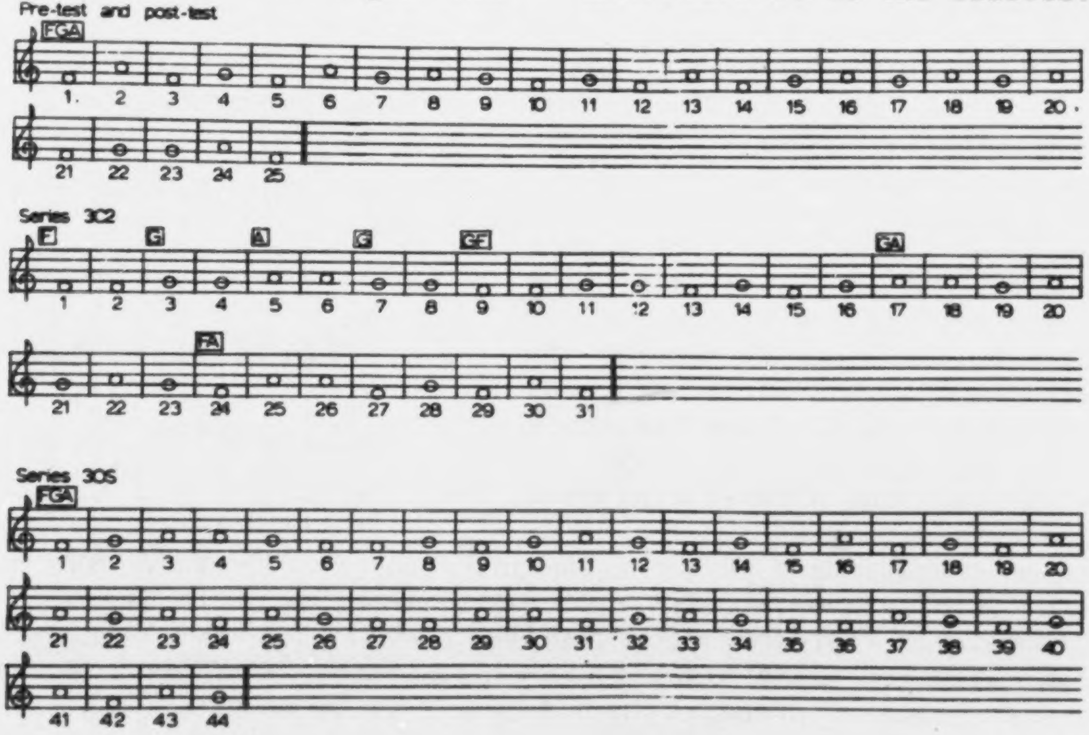

Pre-test and post-test
FGA
1. 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20.
21 22 23 24 25
Series 3C2
F
G
A
G
GF
GA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
FA
21 22 23 24 25 26 27 28 29 30 31
Series 3OS
FGA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
41 42 43 44

Fig.10 Order of tonal stimuli presentations for Study III

of the program might be obtained. For the pretest, the three key choices were illuminated and the subject was told to find the key which played each tone heard.

The training procedure employed as the mode of gradual progressions of learning steps, the presentation of the stimuli in one, two, and three key-tone matching contexts. The series with one and two matching contexts is shown in Figure 10 as series 3C2. In this series, items 1 through 8 were single match trials; only the appropriate key cue was given in each case. Items 9 through 16 were successive matching trials with the G and F key-tone associations; items 17 through 23, with the G and A key-tone associations; items 24 through 31, with the F and A associations. Thus, series 3C2 provided training in responding to each of the three key-tone associations separately, and to the three combinations of two tone contexts, with only the appropriate keys in each context cued. Series 3C2 was followed immediately by series 30S (also shown in Figure 10), which consisted of a 44-item ordered sequence of the three key-tone associations.

Results and Discussion. Figure 11 shows the cumulative error records of S44 and S46, two new first grade subjects on this program. Both subjects produced chance performances in the pretest: S44 made 14 errors; S46, 16. In 3C2, S44 made 2 errors; S46, none; in 30S, S44 made 8 errors; S46, 6. In the posttest, S44 made 5 errors; S46, 2. In terms of pretest, posttest score differences, both subjects showed "marked improvement." However, more than "improvement" as such, the major concern of the study was the development of a program, the performance of which would carry subjects to the desired terminal behavior in a clear and efficient manner. The increase in error rates in 30S, however, opened to question the relationship between program and improvement. Just how effective was the present procedure in teaching the three tone matching task? Could the procedure be improved to eliminate the increase in errors in the three tone matching context?

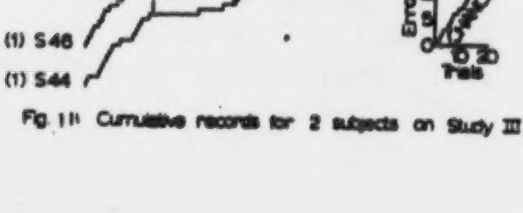


Fig. 11: Cumulative records for 2 subjects on Study III.

The control experiment explores these questions. It was based on the present programming format and employed an essentially similar mode of gradual progression. Mainly altered were feedback and reinforcement variables. Since the appropriateness of musical tasks and the effectiveness of teaching procedures represent problem areas in adult as well as in youth musical education, the control experiment was conducted with individuals in both categories.

## CONTROL EXPERIMENT

Subjects. Fourteen youth subjects from the same school population were used in this experiment. All but one were new subjects. The subjects were selected on the basis of control experiment pretest scores; the new subjects produced scores below 40 percent correct. The subject who participated in some earlier work produced a pretest score of 53 percent correct.

Fifteen adult subjects participated in this experiment, with ages ranging from 24 to 78 years. Adults were selected on the basis of their claims to (a) not being "musical," (b) not being able to read music, and (c) not having any recent or prolonged music study.

Preliminary Procedures for Youth Subjects. For the control experiment, it was decided to equate the youth subjects with respect to conceptual skills of "same and different" with visual stimuli, and ability to differentiate between two tones, one half tone apart. This procedure was not followed with the adults. It was felt that giving adults the simple visual stimuli was unnecessary, and giving them the auditory test unwise, since the purpose given for their participation was to help in determining the interest values of a new method of teaching music.

The procedure consisted of three steps. Step 1 was administered on one day; steps 2 and 3, on the following day.

Step 1. Visual materials for step 1, consisted of twelve 5 x 5-inch cards with "same and different" pictures on them. With E and subjects (taken individually) seated facing one another, E presented card 1 and said, "See these two flowers? They are just alike, aren't they? Just the same." E then showed card 2. "But these two are not alike (pointing). One is black and one is white." E then showed card 3, saying, "Now look at these two. Are they alike? Are they the same?" and waited for subject's answer. For each successive card presented, E asked the same questions and waited for subject's answer.

Step 2. Materials for step 2 consisted of two 8-1/2 x 11-inch cards; on one card was centered a 2-1/2-inch diameter black circle, on the other a 5 x 5-inch white square. For this step, the keyboard was fitted with two 2 x 2-inch wooden blocks, one block placed at each end of the keyboard. On the left block was a small card with two black dots on it; this was the "same" block. On the right block, the "different" block was a card with a small dot and a white square. E seated the subject at the keyboard and siad, "do you remember the game we played yesterday? Well, today we're going to play one that's a little bit different. Today we're going to play one with pictures and with these two blocks (indicating the two blocks on the keyboard). What do you see on the top of this block (pointing to the "same" block)? Are they the same, just alike? What do you see on the top of this block (pointing to the "different" block)? Are these the same, just alike? Now I will show you two pictures, first one and then the other. If the two pictures I show you are the same, you push the block that has the same pictures on it. If the two pictures I show you are not the same, you push the block that has the pictures that are not the same on it." E stood beside the keyboard console, held the two cards behind it, lifted first one and then the other so that the cards were in the subject's line of vision, and asked the subject to push the appropriate same or different block to correspond to the two pictures that were shown. E presented approximately five to seven same-different card combinations to each subject.

Step 3. After making sure that Ss pushed the appropriate blocks in step 2, E said, "Now, instead of looking at pictures we're going to listen to sounds. You are going to hear two sounds, just like you saw pictures. If the two sounds are the same, just alike, you push the block with the pictures that are just the same. If the two sounds are not the same, not alike, you push the block with the pictures that are not the same. Are you ready to listen?"

The tonal stimuli for this phase of the procedure were C followed by C as the "same" pair, and C followed by C# as the "different" pair. These "same" and "different" stimuli were presented in a 30-trial sequence following the Gellerman Series (Gellerman, 1933). After the first three trials were presented, E said, "Good. I'm going to go and sit in that chair (indicating a chair behind the console at one end of the room). You can finish this game by yourself because you know how to do it. Try to push the right block every time." And the remaining 27 trials were presented.

## Control Experiment Procedures

Pretest (All Subjects). Pretest (and posttest) consisted of a 30-item randomized sequence of the "do, re, mi" configuration from 3RS, transposed to the F, G, A key-tone associations. Pretest (and posttest) were given in an extinction procedure, that is, all responses were made on silenced keys; subjects received no differential feedback (as to whether or not their responses were correct). This extinction procedure was adopted as a warranty against the possibility of learning taking place in the pretest.

With subject seated at the keyboard, E said, "For the game we're going to play now, we are going to use some of these keys. Usually, each of these keys makes a sound when you push it down. (E demonstrated by pushing down keys which "played.") We're not going to play with all of these keys, just some of them. Something special is going to happen up here (pointing to the row of cue lights over the keys) to tell us which keys we're going to use. (Cue lights over the F, G, and A keys were illuminated.) Can you guess what it is that tells us which keys we're going to use? Which key does this light tell us to use? This light? And this one? (E pointed to each of the keys lighted.) Good. These are the three keys we're going to use. Remember, I told you that usually these keys make a sound when you push them down. For this game we're going to turn off the sounds so that the keys will not play. (E demonstrated by pushing down on the F, G, and A keys which were now silenced.) But the machine will make sounds. It will make the sounds that go with these three keys. After each sound that the machine makes, you guess which key goes with that sound and push it down. All right? Are you ready to listen?" Youth subjects were given the pretest on one day, the training and posttest on the following day.

Training procedures. With the three choices available (illuminated), training included presentations of the tonal stimuli in 1, 2, and 3 matching context segments in the following order: (1) tone F; (2) tone G; (3) tones F and G; (4) tone A; (5) tones F and A; (6) tones F, G, and A. In the two and three-tone contexts, the stimuli were presented in random order. Progression through training segments 1-5 was based on subjects meeting a criterion contingency of five correct responses in succession in each segment. The duration of each tonal presentation was 1.5 sec.

Three procedures for reinforcement and feedback were used in the training.

Procedure 1. Correct responses were reinforced by the reproduction of sample tones, plus an illuminated lamp signal; incorrect choice keys in each trial were silenced, thus incorrect responses were made in an extinction procedure.

Instructions: E, "For this game, we're going to use the same three keys that we used in the game we played the last time. Do you remember which keys they were? Good.

These three keys are the ones we will use. Now, when the game starts, the machine is going to play a sound for you. You can make the same sound come again by pressing down on one of these keys. So, your part in the game is to learn which one of these keys will make the sound that the machine made, come back again. You will be able to tell when you have pushed the right key because, first, you will hear the same sound come back again, and this red light (pointing) will go on. If you press a key that doesn't make a sound, just wait for the next sound and try to get that one right.

Subjects

Youth subjects:

S170, S171, S148, S60 from kindergarten, and S184 from the first grade.

Adult subjects:

AS3, AS6, AS7, AS9, AS1.

Procedure 2. Correct responses were reinforced by reproduction of sample tone solely (no light); incorrect responses were made in the extinction process as above (no buzzer).

Instructions: E, "For this game we're going to use the same three keys that we used in the game we played the last time you were here. Do you remember which keys they were? Good. These are the three keys we are going to use. Now, when the game starts, the machine is going to play a sound for you. You can make that same sound come again by pressing down on one of these keys. So, your part in the game is to learn which one of these keys will make the sound that the machine plays, come back again. You will be able to tell when you have pushed the right key because you will hear the same sound come back again. If you press a key that doesn't make a sound, just wait for the next sound and try to get that one right."

Subjects

Youth subjects:

S168, S169 from the kindergarten, S26 from the first grade, and S198 from the second grade.

Adult subjects:

AS10, AS12, AS15, AS16, AS17.

Procedure 3. In this procedure, the three choices were "live"; correct and incorrect responses were differentially reinforced by reproduction of associated tones of each choice.

Instructions: E, "For this game, we are going to use the same three keys that we used in the game we played the last time you were here. Do you remember which keys they were? Good. These are the three keys we're going to use. Now, when the game starts, the machine is going to play a sound for you. You can make that same sound come again by pressing down on one of these keys. So, your part in the game is to learn which one of these keys will make the sound that the machine plays, come back again. You will be able to tell when you have pushed the right key because you will hear the same sound come back again. If you press a key that doesn't make the same sound come again, just wait for the next sound that the machine makes, and try to get that one right."

Subjects

Youth subjects:

S142, S146, S166 from the kindergarten; S63 and S174 from the first grade.

Adult subjects:

AS18, AS19, AS21, AS22, AS23.

Following instructions, subjects were given the training program, as outlined earlier, under one of these three procedures. The control experiment was administered to the adult subjects (individually) in one session. In the instructions to adults, references to past sessions were omitted.

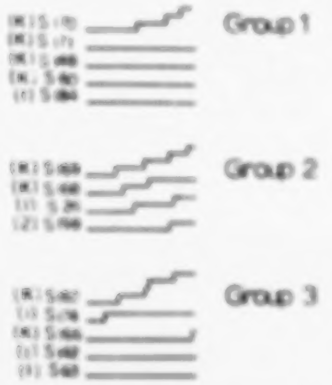
Posttest. The posttest immediately followed the training. The following instructions were given for the posttest. E,"Good. Now let's play the game again where only the machine makes sounds. The keys will not sound (and the red light will not go on). After each sound the machine makes, you press down on the one key that you learned goes with that sound."

Procedural note. Since the programmer unit was designed to present the tonal stimuli in a fixed, prearranged order, it could not handle the criterion contingencies required in this experiment. The apparatus was altered, therefore, so that the stimuli could be presented by manual operation from a small room adjoining the experimental room.This manual presenter was operated by the writer. A one-way mirror placed in the connecting door permitted the writer to observe subjects working at the keyboard response unit.

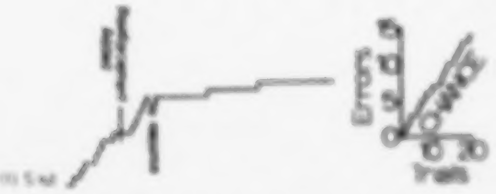
This procedure provided the required flexibility in programming, but the substitution of a human programmer presented problems. An automated programmer would have presented the stimuli in the required order,evaluated and counted the responses, and, at the appropriate times, moved on to the next program segment, all without having to observe the subject and without being diverted or otherwise influenced by the behavior occurring in the other room. Performing all these chores accurately and objectively was not an easy task, at least for this writer.

As a result of the "human element," variations in length of training segments and, in the cases of a few adults, omissions of segments, occurred. However, other aspects of the study were not violated. The general procedure was maintained; the manual device assured uniformity in the duration of stimuli presentations; the writer did not enter the situation physically or verbally; objective records of the proceedings were obtained.

Results and Discussion. In the preliminary procedure for youth subjects, no subject made errors with the visual stimuli in Steps 1 and 2. Figure 12A shows the records on the "same-different" tonal step (Step 3) of the 14 youth subjects who were retained for the experiment proper. This task presented no difficulties for these subjects. Performances were perfect, or with but few scattered errors, which may have been due in part to intrusions of extraneous stimuli (the experimental room was not soundproofed).



A. Performances on step 3 (Same-Different discrimination): no feedback (extinction procedure)

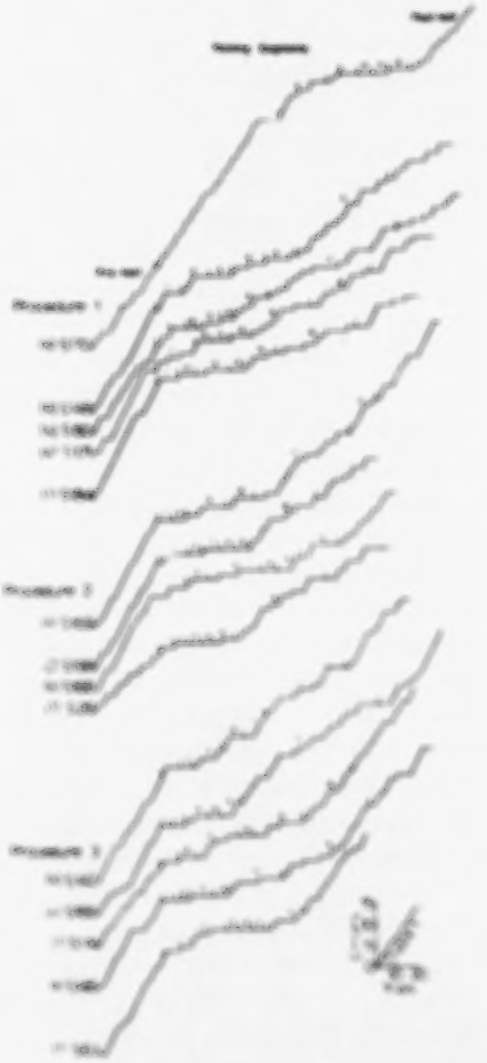


B. Record of S163, showing effect of reinforcement.

Fig 12 Cumulative records of performances on step 3, preliminary procedures, Control Experiment.

One subject, S163, completely failed to make the "same-different" discrimination in the first 15 trials, and thus presented an interesting problem, namely, could he be taught the distinction by nonverbal, objective means? It was decided to remove this subject from the experiment, but to follow up immediately on the problem. The record of the proceedings with S163 is shown in Figure 12B. At item 16, conditions were arranged so that the red lamp signal would illuminate on correct responses. The subject was told by E, "As you play the game, try to make this light (pointing to red lamp) go on every time." Note that in items 1 through 15 S had been responding to the "same" block. He continued to do so in items 16, 17, 18, which, being correct, activated the lamp signal. Items 19 through 23 were "different" trials, but the subject responded incorrectly and the lamp signal did not illuminate. Subject did not respond in item 24. In item 25, subject was heard to say, "They are a little bit different," and responded to the correct block. Subject was then given 49 additional "same-different" trials, which he performed with but two errors.

Figure 13 shows the cumulated error records of the youth subjects on the experiment proper. With the possible exception of S171, subjects did not learn the

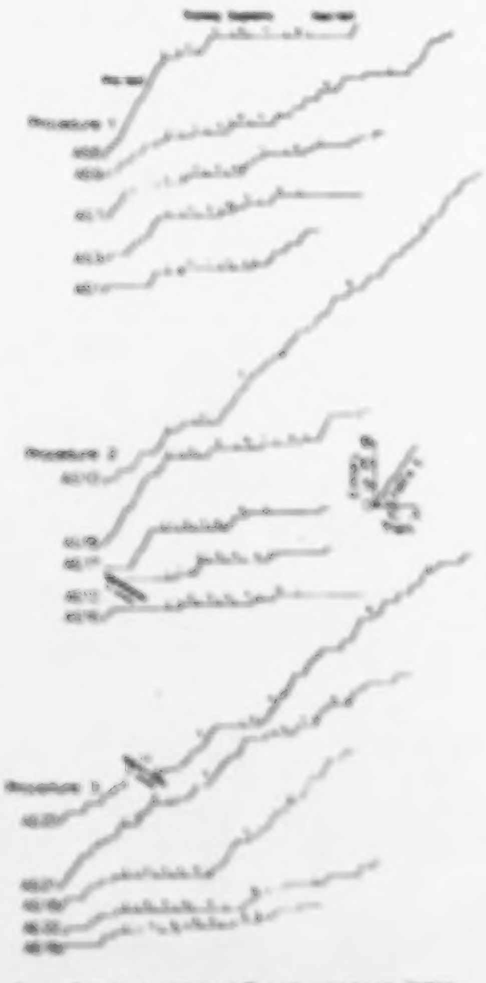


[illegible]

task in the pretest. That S171's curve flattened out in the last few items could have indicated he might have learned. This would have been known only had the test continued for another 10 items or so. However, as was shown in the exploratory studies, several correct responses in a row could have been made for reasons other than "knowing" the specified task.

With the availability of the three choices (illuminated keys) in the training procedures, it was expected that "exploratory" errors, or errors due to adaptation, would occur in the first training segment. It was interesting that such errors were relatively few and that they occurred mainly in Procedure 1. Errors in the one and two match contexts (segments 2-5) were few. Errors increased in the context with three matches (segment 6)preceding the posttest. The greatest increase occurred in Procedure 3. In the posttest, all subjects but one(S174 in Procedure 3) recorded gains in the reduction of errors. The greatest gain occurred in Procedure 1; the least, in Procedure 3.

Figure 14 shows the cumulated error records of the adult subjects. Adults totaled 50 percent fewer pretest errors than the youth subjects, with individual error rates ranging from 3 to 23. Errors increased in segment 3 for AS10 and AS21; errors increased in



[illegible]

segments 3 and 5 for AS23. Eleven subjects (AS6, 7, 3, 1, 15, 17, 12, 16, 21, 22, 19) successfully performed in the context with three matches preceding the posttest: four subjects (AS9, 10, 23, 18) produced sufficient errors to indicate difficulty in the segment. Reductions in posttest errors were made by ten subjects (AS6, 7, 3, 15, 17, 12, 16, 23, 21, 19); three subjects (AS1, 10, 18) increased their error rate in the posttest; two subjects (AS9, 22) showed no change in performance. With respect to the gains in reduction of errors in the posttest in the three training procedures, the results paralleled those of the youth subjects.

In general, this study supports the conclusions reported by other musical investigators that (a) pitch discrimination is improvable with training (Gibson, 1953), and (b) the procedures employed did not achieve maximum improvement. What, then, are the requirements for the achievement of greater improvement? Some suggestions which have been made in this regard include the need for "better work methods" (R. H. Seashore, 1935) "right habits of attention" (F. O. Smith, 1914), "a greater variety of methods" (Wyatt, 1945). It is this writer's suggestion that the procedures themselves must be looked into, for when they are effective, they effectively control the subject's "attending" and "working" behavior. A case in point is the example of S163 (Figure 12B). However, only when the effects of procedures are known can they be evaluated.

Although most subjects in this experiment improved in their posttest performances, the instructional program was far from satisfactory; it did not constitute a smooth or dependable series of learning steps to the terminal behavior. No youth subject exceeded 85 percent on the posttest; of the adults, seven exceeded 85 percent.

In the present single tone matching procedure, the question arises as to the role of the "absolute" properties (specific frequencies), or the "relative" properties("distances" between frequencies) of the stimuli (since a tone in a series,in addition to its individual frequency, possesses an intervallic relationship with the preceding tone[s]).

In this study, cues for either "absolute" or "relative" judgments were not isolated or independently manipulated. The use of the single tone as the stimulus unit would seem to emphasize the "absolute" cue. On the other hand, the brief period between trials would seem to operate in favor of the intervallic (inter-trial) relationship.

Just how much time must elapse between trials for"absolute pitch" judgments is an open question. Opinions have ranged from 20 seconds (Lundin, 1963) to one trial per day administered before subject had the opportunity to hear any music (Petran, 1932). By their comments to the effect that they "caught on to the logic" or "to the system" of the task, several adults indicated they were using "relative" cues. It is true that sounds of definite pitch are the raw materials of music. It is also the case that most musical conventions and cues are based on intervallic relationships rather than on absolute frequency values of individual tones. It was possible that for most subjects, neither set of cues was present in sufficient strength to matter.

In retrospect,correct responding could have been achieved in segments 1 to 5 without having to attend to either set of cues. In the single match contexts in Procedures 1 and 2, the subject had only to respond to the illuminated key that "played" a tone; no tonal discrimination was required. In Procedure 3, the subject could have differentiated between the trial tone and "other" tones. In contexts with two matches, correct responding could have been achieved on the basis of "same - different" cues, without having to attend to specific frequencies or to specific intervallic relationships. If such were the case, then segments 1 to 5 would constitute inadequate training for the three-tone discrimination, except perhaps for providing opportunities for incidental learning.

Based on the foregoing, the difficulty of the terminal task cannot be attributed to the addition of the third key-tone association, or merely to the inadequate use of "absolute" or "relative" cues. Ultimately, the problem must be the understanding of the behavioral process, which is set off by the discriminative stimulus as the occasion for the response, e.g., for the desired behavior to occur. A model of the present key-tone matching task might be represented as follows: the subject is (1) presented with a discriminative stimulus (tone), (2) which is no longer present at the time of the response, (3) which occurs along a different stimulus continuum (keyboard), (4) which produces a tonal consequence, (5) which he must recognize as identical with the absent tone (tonal memory behavior), (6) which must also reinforce the key press response (position to key along the keyboard continuum). After this series of events, the process begins all over again when the subject is presented (1) a current stimulus, (2) which he must now differentiate from the stimulus of the preceding trial, and so on. Under conditions which include three stimuli each along two different continua, and in which samples and matches are not simultaneously present, simply "keeping the records or places straight" represents behavior by no means simple. If this represents an accurate model, it is easy to see that natural contingencies of reinforcement for "musical" behavior within such a complex process would be quite rare in everyday environments and, indeed, in most instructional procedures.

## SUMMARY

A musical task involving pitch discrimination was developed and taught to children and adults. The task was a single tone matching procedure employing three organ tones and keys in the "do, re, mi" configuration, in which individually presented ("heard") tones, served as stimuli controlling selection of keys for tone-matching responses. The design of the study included an exploratory phase, in which sequences of material were manipulated and behavioral (performance) characteristics investigated, and a control experiment, in which reinforcing stimuli (feedback contingencies) were manipulated. Data describing errors as a function of trials and other relationships between stimuli and response events were collected. These were plotted in the form of individual cumulative records and stimuli-response tracking charts, respectively.

Pre-adult subjects ranging in age from four to eleven years were recruited from an academic and professional population attending a private school. Adult subjects (ranging in age from 24 to 78 years) were selected on the basis of their claims to (a) not being "musical," (b) not being able to read music, (c) not having any recent or prolonged music study. All subjects worked at the keyboard of a specially adapted organ by which tones were presented, responses recorded, and feedback provided.

The study was primarily concerned with the following questions:

1. Would this key-tone matching task be "easy" or "difficult"?
2. Could performance be improved by employing the objective procedures described?
3. What behavioral characteristics could be revealed by the use of such procedures?

Data from the exploratory phase (conducted with youth subjects) indicated that matching "do, re, mi" keys to their associated tones in single tone matching trials was not as "easy" a task as one might have expected. Analyses of stimuli-response tracking charts indicated that the major "pre-solution" behavior patterns were affected by the programmed sequences of material, and by the one, two, three, order inherent in the stimuli themselves. These data also provided evidence of the systematic nature of such performance, and that the relationships between sequence of material and patterns of responding can be subjected to experimental control.

In the control experiment (conducted with youth and adult subjects) three different reinforcement procedures (feedback contingencies) were tested on a sequence of material developed in the exploratory phase. Procedure 1 provided production of the trial tone plus a red light signal for a correct response, and silence (no consequence) for an incorrect response. Procedure 2 was the same as procedure 1, with exception that the red light signal was withheld. In procedure 3, correct and incorrect responses produced the associated tones of the keys pressed. Control experiment data (cumulative error graphs) revealed improvement in 13 out of 14 youth subjects, and in 10 out of 15 adult subjects. The procedures in which subjects heard the correct tones and not the incorrect tones in training (1 and 2) produced more learning than the procedure (3) in which subjects heard the correct as well as the incorrect tones. Also, more learning was achieved under procedure 1, which provided the red light consequence in addition to the tone for correct responding, than under procedure 2, which provided the tone only (no light).

A behavioral model of key-tone matching was suggested as the framework for further research. The implications of the study were seen to extend beyond musical learning theory, to auditory learning theory in general. The relevance of the findings to education in general and to adult education in particular was indicated.

## REFERENCES

Cameron, E. H., and Steele, W. M. Effects of practice in the discrimination and singing of tones. *Psychol. Monogr.*, 1917, 23, No. 3.

Gellerman, L. W. Chance orders of alternating stimuli in visual discrimination experiments. *J. genet. Psychol.*, 1933, 42, 207-208.

Gibson, Eleanor. Improvement in perceptual judgments as a function of controlled practice and training. *Psychol. Bull.*, 1953, 50, 401-431.

Hilgard, E. R., and Marquis, D. G. *Conditioning and learning*. 2nd ed. revised by Gregory A. Kimble. New York: Appleton-Century-Crofts, Inc., 1961.

Hively, W. Programming stimuli in matching-to-sample. *J. exp. Anal. Behav.*, July, 1962, Vol. 5, No. 3, 279-298.

Holland, J. G. Teaching machines: An application of principles from the laboratory. *J. exp. Anal. Behav.*, October, 1960, Vol. 3, No. 4, 275-287.

Krechevsky, I. 'Hypothesis' in rats. *Psychological Revue*, 1932, 39, 516-532.

Lumsdaine, A. A., and Glaser, R., (Eds.). *Teaching machines and programmed learning: A source book*. Washington, D. C.: National Education Association of the United States, 1960.

Lundin, R. W. Can perfect pitch be learned? *Music Educators Journal*, April-May, 1963. Vol. 49, No. 5, 49-51.

Mursell, J. L. Growth processes in music education. In N. B. Henry (Ed.), *Basic concepts in music education*. Chicago: The National Society for the Study of Education, 1958, pp. 140-162.

Mursell, J. L. *The psychology of music*. New York: W. W. Norton & Company, Inc., 1937.

Ortmann, O. *Problems in the elements of ear dictation*. Research Study in Music, No. 2. Baltimore: Peabody Conservatory of Music, 1934.

Petran, L. A. An experimental study of pitch recognition. *Psychol*. *Monogr*., 1932, 42, No. 6.

Porter, D. Teaching machines. In A. A. Lumsdaine and R. Glaser (Eds.), *Teaching machines and programmed learning*. Washington, D. C.: National Education Association of the United States, 1960, pp. 206-214.

Seashore, C. E. *Psychology of music*. New York and London: McGraw-Hill Book Company, Inc., 1938.

Seashore, C. E. *The psychology of musical talent*. Boston: Silver, Burdett and Company, 1919.

Seashore, R. H. Improvability of pitch discrimination. *Psychol*. *Bull*., 1935, 32, 545.

Skinner, B. F. The science of learning and the art of teaching. In *Cumulative Record*. New York: Appleton-Century-Crofts, Inc., 1961, pp. 145-157.

Skinner, B. F. Teaching machines. *Science*, October 24, 1958, Vol. 128, No. 3330, 969-977.

Skinner, B. F. Why we need teaching machines. *Harvard* *educ*. *Rev*., Fall, 1961, Vol. 31, No. 4, 377-398.

Smith, F. O. *The effect of training in pitch discrimination*. Iowa City: University of Iowa, Studies in Psychology, 1914, VI.

Wolner, M., and Pyle, W. H. An experiment in individual training of pitch deficient children. *J*. *educ*. *Psychol*., 1933, 24, 602-608.

Wyatt, Ruth F. Improvability of pitch discrimination. *Psychol*. *Monogr*., 1945, 58, No. 2.

# TEACHING A SKILL OF DATA INTERPRETATION BY PROGRAMMED INSTRUCTION

Vernon W. Stone

## INTRODUCTION

Many experiments have been conducted employing programmed instruction in studying sundry phases of the teaching-learning process. All in all, notable success has been attained as a result thereof. To be sure, several self-instructional programs currently on the market deal with the direct teaching of theoretical statistics and probability. In such programs, the subject innovative methodology is the focus. Pursuant to the supposed acquisition of the defined skill, however, the learner is expected to perform in an exacting quantitative manner.

One unit of theoretical statistics--simple correlation, or relationship--has been selected for the subject experiment. Emphasis has been placed on judging from scatter diagrams--without learning formal statistical concepts or computational procedures--the magnitude, direction, and linearity of relationships between variables.

Hence, the criterion behavior, or the final outcome of the instruction, may be stated as follows: *to determine intuitively--by non-computational inspection--selected relationship indices of a bivariate scatter-diagram*. The intuitive response of the learner is clearly in contradistinction to the usual quantitative approach, outcome, and practice.

Although this experiment has been predicated upon methodological and content innovations, there is also a practical aspect involved: countless individuals constantly demonstrate the need for determining visibly observable relationships. Hence, the interpretative need is suitably attended through this approach.

## METHODS AND PROCEDURES

*The Instrument.* The instrument employed was a teaching - learning program especially constructed for conducting this experiment.[1] The program has a total of eighty-seven (87) frames, divided among six (6) sets as follows: Set A, Relationship, 16 frames; Set B, The Geometric field, 14 frames; Set C, Bivariates, 18 frames; Set D, Qualitative Magnitude, 17 frames; Set E, Directionality, 10 frames; Set F, Regression, 12 frames. (See Fig. 1 for examples.)

The entire program --including including instructions, the pretest, the program proper, and the post-test--requires approximately 128 minutes ( 2 hours, 8 minutes ) per subject.

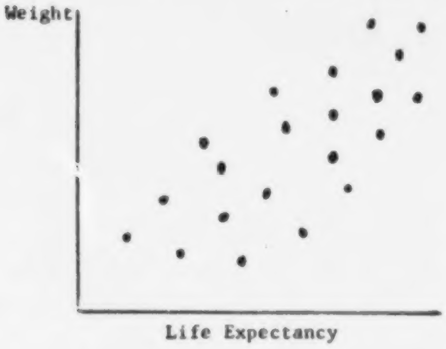
An exhibit is presented at the beginning of each of the six sets. Basic, conceptual data are con-

---

[1] The program is entitled "A Self - Instructional Program for Determining, by Non - Computational Inspection, Selected Relationship Indices of a Bivariate Scatter-Diagram." This program--along with the supplemental materials in connection therewith -- is on file in the library of the Committee on Programmed Instruction, Harvard University, Cambridge, Massachusetts.

Set D: QUALITATIVE MAGNITUDE

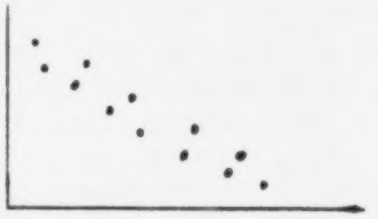
D-17. The relationship between the bivariates, weight and life expectancy, is (1) ______, which means that weight probably has some influence on one's (2) ______ ______.



(1) Moderate (2) Life Expectancy

Set E: DIRECTIONALITY

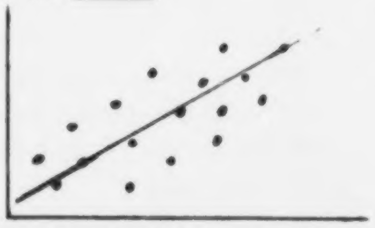
E-10. The magnitude of the relationship in the figure is (1)______, and the directionality is (2) ______.



(1) high (2) negative

Set F: REGRESSION

F-1. The estimated magnitude of the relationship of the two variables is (1) ______; the directionality is (2) ______; and the regression is (3) ______.



(1) Moderate (2) Positive (3) Linear (Rectilinear)

tained therein. In each case, the set exhibit is read and/or studied prior to the learner's responding to the questions of the given set. Then, during answering items, the exhibit is referred to as often as is required. These illustrations are labelled as exhibits, and they are numbered consecutively numerically for a given set. Within a set, some questions are based upon individual question illustrations. In such instances, these question illustrations are labelled as figures,and they are numbered consecutively numerically for a given set.

The Subjects. Ten (10) Ss--five men and five women--were selected on the basis of a pre-test.

The pre-test and post-test are precisely the same, except for the fact that the scatter-diagrams are arranged in two different orders. The same test was used for the purpose of façilitating a comparison between the two performances --before and after exposure to the teaching - learning program. Ten (10) scatter-diagrams -- selected from the figures contained in the program comprise one of the tests. Each test figure requires three sub-responses--on qualitative magnitude, directionality,and regression. The items were presented to the Ss in booklet form, in such a way that only one diagram, or figure, was visible at any one time. Moreover, the S was unable to review previously decided responses.

The pre-test, which served as a suitable factor in the selection of Ss, was administered individually immediately upon acquiring a prospective S. Applicants who passed the test by answering at least five of the three-part questions correctly were eliminated. The ten Ss who were accepted for the project averaged 8.7 errors on the ten pre-test questions.

On the basis of the subsequent post - test performance, upper and lower criterion groups of the Ss, comprising the upper 50 per cent and the lower 50 per cent of the Ss, respectively, according to the post-test score distribution, were formed.

The five Ss who fared best on the post-test, the upper group, showed a pre-test mean score of 47.4. The lower group showed a pre-test mean score of 52.0. The overall mean score on the pre-test, for all ten Ss, was 49.6.

## RESULTS

Post-Test. (Criterion Behavior) The post-test, which served as an inventory of programmed learning, was administered at the very end of the programmed instruction.

The ten Ss averaged 1.4 errors on the ten questions.

It must be emphasized here that two scoring methods were used in scoring both the pre-test and the post-test:

(1) The entire answer was considered incorrect if one or two of the three parts were incorrect; and

(2) Three points were deducted for each incorrect part response, thus permitting the S a 10-point advantage.

The former method was used in scoring the pre-test for participation or elimination purposes, on the theory that a prospective S could guess the correct responses for one or two of the question parts. On the other hand, it was reasoned that one's ability to respond correctly to all three parts of the question was more attributable to his knowledge of same. The latter scoring method--giving credit for correct part responses--is the one employed in the before-and-after comparison.

The five Ss in the upper group showed a mean post - test score of 98.2, while the five Ss in the lower group showed a mean post - test score of 91.0. The overall mean post - test score, for all ten Ss, was 94.6.

The reliability of the post-test was computed by four different approaches. Needless to say, the reliability coefficients were disappointing. The highest coefficient obtained was .72 ± .16 -- an odd-even computation scored on the basis of ten responses rather than thirty responses. Perhaps the small number of Ss, the small number of post - test items, and the accidental arrangement of items account, in large part, for the low reliability factor. When an instrument correlates so very poorly with itself, there is some question as to whether or not it should be employed as an external criterion variable. In brief, when an instrument fails to correlate internally, it can scarcely be expected to correlate externally.

Pertinent Statistics. As indicated earlier, the use of the pre-test and post-test enabled the experimenter to compare the results on a more definitive basis. Since the t value shown is 15.66, there is no question but that the performances of the ten Ss on the two tests are different. Accordingly, there is an exceedingly strong indication that the auto - instructional program played its designed role.

The pertinent statistics are as follows:

| Statistic | Pre-Test | Post-Test | Overall |
|---|---|---|---|
| N | 10 | 10 | |
| M | 49.6 | 94.6 | |
| 6 | 7.32 | 4.20 | |
| 6M | 2.44 | 1.40 | |
| 6dM | | | 2.81 |
| dM | | | 45.0 |
| t | | | 15.66 |
| P | | | >.01 |

## DISCUSSION

Error Analysis. As previously indicated, the participants averaged 8.7 errors on the ten pre-test questions. Also, the ten Ss averaged 1.4 errors on the ten post-test questions.

With respect to the program proper, the ten Ss had an average error range from .5 of an error on 2 of the 10 frames of Set E to 2.1 errors on 13 of the 18 frames of Set C. The average number of errors per set was directly proportional to the number of frames in a set. The ten Ss averaged, on the total program, 7.2 errors on 42 of the 87 frames.

Inter-Correlation of Variables. The Pearsonian product-moment coefficient of correlation was computed for fourteen (14) variables as follows: each of the individual sets, A through F, 6; combined sets, A, B, and C, 1; combined sets, D,E, and F,1; total program, 1; pre-test, 1; post-test (each of three parts scored separately; then, the total test), 4.

In order to facilitate computations, all variables were consistently and uniformly scored on a 10-point basis, thus rounding the score to the nearest whole number. Accordingly, it was necessary to deal with only ten different numbers in computing the coefficients of correlation; that is to say, only ten different numbers were necessary insofar as the raw scores of the variables were concerned.

## SUMMARY AND CONCLUSIONS

Certain strengths of the program are to be noted. In the first place, it was demonstrated that the program actually taught what was designed to be taught--on the basis of the significance test applied to the pre-test/post-test performances. Thus, a program designed to do so can establish a skill in judging statistical relationships from plots of data. Moreover, overall, the comments by the Ss were particularly gratifying.

# ATTEMPTS TO PROGRAM MATERIALS TO TEACH CERTAIN ELEMENTS OF A PHONIC REPERTOIRE

**Wells Hively and Helen M. Popp**

## INTRODUCTION

A "phonic repertoire" is a *verbal repertoire* (Skinner, 1957, pp. 65-71) based on phoneme-grapheme correspondences. When phonemes are stimuli and graphemes occur as responses, such a repertoire is *transcriptive*. When graphemes are the stimuli and phonemes occur as responses, the repertoire is *textual*. Transcriptive phonic repertoires have to do with Spelling, and textual phonic repertoires have to do with Reading.

Research in the area of Elementary Reading is commonly concerned with comparing and evaluating relatively elaborate teaching "systems" and not with analyzing their sub-components. Therefore, very little is known about the characteristics of an effective phonic repertoire, about how it may be most effectively taught, or about what role it may best play in the complex task of learning to read. The following exploratory experiments represent the beginnings of an attack on one of these problems: how a phonic repertoire may be most effectively taught.

Until some progress has been made in analyzing methods of teaching regular phoneme-grapheme correspondences, we have chosen to ignore the problem of teaching irregular or inconsistent correspondences. We have limited ourselves, for the present, to portions of written English in which there is a one-to-one correspondence between graphemes and phonemes.

The acoustical properties of individual phonemes vary considerably, depending upon the other phonemes that precede and follow them in particular utterances. A child must therefore be able to "respond to" a given phoneme-grapheme correspondence in a variety of contexts if he is to be said to "know" it. Children may be taught the sounds of individual letters or letter groups "in isolation," later combining them to form words (see Eaton and James, 1956), or they may be led to induce phonic correspondences by contrasting words taught through a "whole word" method (see Hildreth, 1958). In the latter situation, words are ordinarily selected on the basis of semantic and grammatical considerations, and not in order to provide an optimum set of phonic contrasts. However, a few authors of reading materials have introduced phonic elements systematically (cf. Daniels & Diack, 1959), and in the materials by Bloomfield and Barnhardt (1961) semantic considerations are subordinated to sequences of phonic contrasts.

In so far as we know, no studies have been made of the effect of various alternative programs of word contrasts in shaping a given phonic repertoire, or of the various ways in which specific phoneme-grapheme correspondences may be "isolated" and subsequently combined. Controlled experiments of this sort are difficult to carry out without special instrumentation, and virtually impossible in the classroom.

[Reprinted from The JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR, Vol. 7, No. 5, September 1964.]

## CRITERION BEHAVIOR

In the following exploratory experiments children were required to select, from among a set of alternatives, a grapheme which corresponded to a given phoneme. A shorthand way of referring to this behavior is to call it *matching graphemes to phonemes*. The phonemes were either /s/ or /l/ and the graphemes were the lower case letters *l* and *s*. These appeared as initial consonants in pairs of monosyllabic words which were otherwise identical in both spelling and pronunciation (e.g., lack, sack; let, set). The study used all 25 such rhyming pairs found among the 6,000 most frequent words (Thorndike & Lorge, 1944).

We attempted to teach children to discriminate between the members of these pairs of written words on the basis of their pronunciation. In each pair this discrimination must be based upon the phoneme-grapheme correspondence of the initial consonants (although in each pair, discrimination of these elements may be based in part upon the context in which they appear). We wished to see whether children could be taught to abstract the phonic relationships, the test of the abstraction being errorless responding to new pairs of words. The following experiments are all concerned with attempts to program materials to teach this behavior.

## EXPERIMENT I: MATCHING GRAPHEMES TO PHONEMES IN THE CONTEXT OF "RHYMING" WORDS

### Apparatus and Procedure

Pairs of words were typed, photographed, and projected onto windows in a discrimination apparatus (Hively, 1962 [described in an earlier chapter]). (The words were typed with an IBM *Executive* typewriter in *Documentary* type and projected twice their original size.) A spoken word was presented via a loudspeaker (mounted above the response panel of the discrimination apparatus) immediately after the written words appeared. The child then selected one of the written words by touching the window on which it was projected. If his choice was correct, the projector changed slides and the next item was presented. If it was incorrect, nothing happened and he could respond again immediately. Spoken words were read from a list by an experimenter in an adjoining room. Another person brought the children one by one to the experimental room and conducted the experiments.

Each child was first given a pre-test. The experimenter showed him pairs of words printed on cards, pronounced one of the words, and asked him to point it out.

The child was then given a visual discrimination test in the visual discrimination apparatus. One printed word was shown in the sample window while this word and its "rhyming" pair was shown in the "choice" window.

All children had previously been trained in a visual matching task, (Hively, 1962) and knew how to operate the apparatus.

The child was next given an auditory discrimination test. Pairs of words were presented over the loudspeaker and the child was asked to judge whether they were the "same" or "different" by pressing designated windows of the discrimination apparatus.

Finally the child was given a *training program*, and the pre-test was repeated as a *post-test*.

### Subjects and Test Results

Subjects were first grade children (in a school system where there was no public kindergarten) tested in the fall, and kindergarten children (in another school system) tested during the spring. The tests were given in order to select children for the training programs. Test data and selection procedures are presented below.

*Pre-test*. A distribution of scores on the pre-tests is shown in Figure 1. Cutting points (arbitrarily chosen at the outset) are shown in the figure. Subjects scoring at or above these points were excluded from further experimentation.

*Visual discrimination test*. Of 46 children who were given this test, seven (15%) failed to reach a criterion of six consecutive correct responses within the 16 items. These subjects were also excluded from further experimentation.

*Auditory discrimination test*. Of 43 children who were given this test, three (8%) failed to reach a criterion of six consecutive correct responses within the 19 items. These subjects were excluded from further experimentation as well.

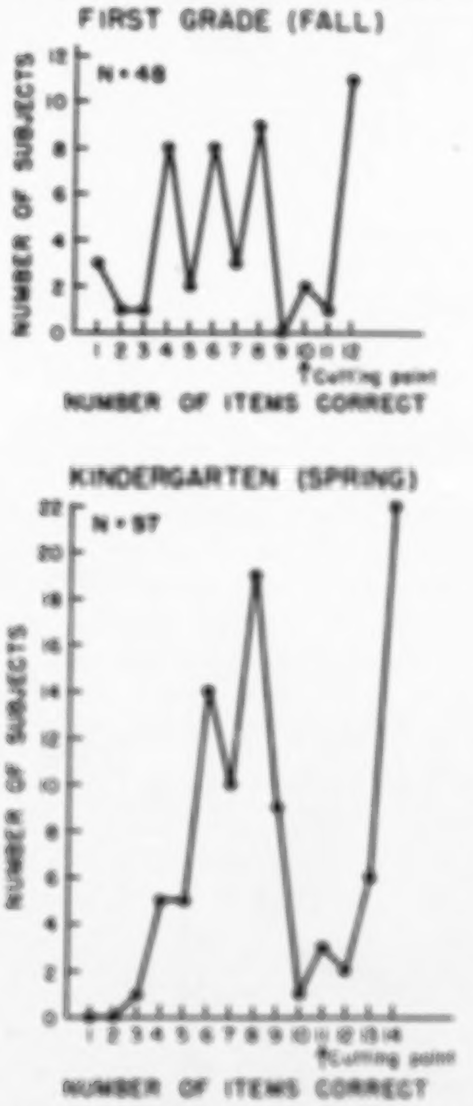


Fig. 1. Distribution of errors on pre-tests.

Training Programs

*Successive pairs*. On each "trial", the same two written words appeared ( in random positions ) and one or the other of them was pronounced (in a random order) until a criterion of six consecutive correct responses was reached. Then another pair of words was presented, and so on through eight successive pairs. Four of the eight pairs, together with four new pairs, were then presented in scrambled order as a test of the phonic abstraction.

Subjects were given the following instruction: "The machine will go if you press the word that you hear. I'll show you how to do it. Listen." *E* responded to the first four items. "Now you do it." After *S* had responded to two items *E* said, "Try to push the right one every time," and retired to a chair behind the subject. The apparatus presented the next set of materials automatically each time criterion was reached. Subjects worked until they completed the program or until they had responded to forty items without reaching criterion (in which case the apparatus automatically shut off.)

The performances of the 15 children who worked on this program are shown in Figure 2. Of the nine subjects who did not complete the program, eight failed to reach criterion on one of the first two pairs. Two subjects who completed all eight pairs failed to abstract the phoneme-grapheme correspondence ( as shown by their test scores ). The pre-test and post-test scores of the remaining four subjects who completed the program satisfactorily suggest that they may have

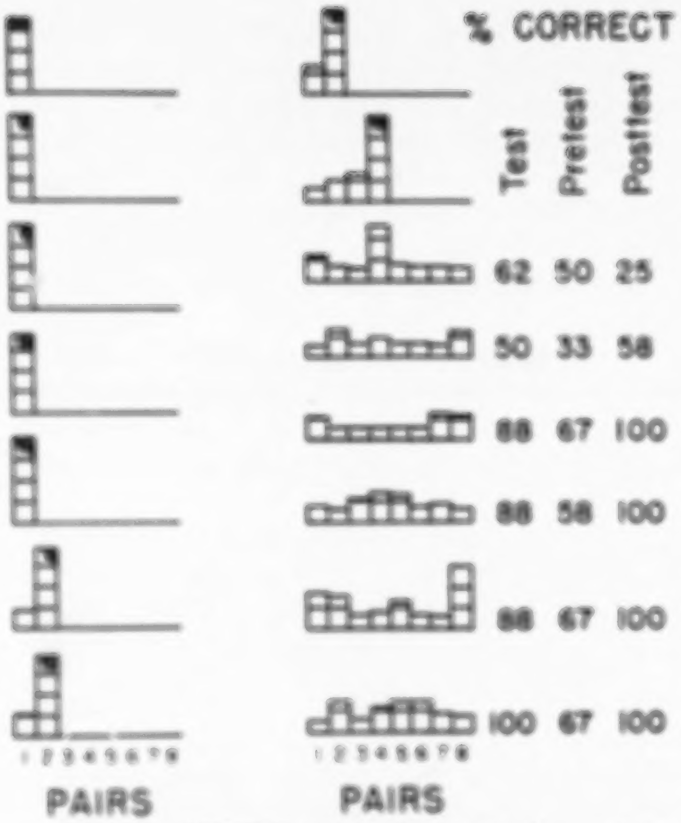


Fig. 2. Performances of subjects on *Successive Pairs*. *Note*: Each graph shows the number of items responded to by each subject in reaching criterion on each successive pair of words. Bars are divided into units representing two items. A broken bar shows where the experiment was terminated. Order of pairs: 1) light-eight, 2) luck-suck, 3) lee-see, 4) lock-sock, 5) lad-sad, 6) lit-sit, 7) let-set, 8) lake-sake. Test: *sad*-lad, *sight*-light, sock-*lock*, lane-*sane*, *lift*-sift, sent-*lent*, long-*song*, jock-*sock*.

learned something from it. One might expect to see a decrease in the number of trials to criterion with each successive pair, if the children were learning the phoneme-grapheme correspondence and making use of it, but there was no such trend.

*Mixed pairs*. In this program the same eight pairs of words were presented, but mixed together. Two instances of every pair (using each word once as a sample) were presented in random order, and this set of items was repeated until subjects reached a criterion of six consecutive correct responses or until they responded to 100 items. Those who reached criterion were given the same test used with Program 1.

Of four subjects who worked on this program, three failed to reach criterion. One reached it within 22 items and made no errors on the test.

*Retained pairs*. In this program, subjects responded to one pair of words until they reached criterion. They were then presented with items representing a new pair mixed with an equal number of items from the old pair. (Items from the same pair did not occur more than three times in succession.) When criterion was again reached, items from another new pair were presented, mixed with items from Pairs 1 and 2 (half of the items were instances of the new pair). In subsequent steps, half of the items belonged to each new pair and the remainder were divided among those preceding. After eight pairs had been introduced, the test used in Programs 1 and 2 was presented. Subjects worked until they completed the program or responded to 40 items from a given step.

Figure 3 shows the performances of the five subjects who worked on this program. Two failed to discriminate the first pair, one failed in Step 3, and one in Step 8. One completed the program and performed satisfactorily on the test.

Discussion

Abstracting initial consonant phoneme-grapheme correspondences, by discriminating between written pairs of "rhyming" words on the basis of their pronunciation, proved to be a difficult task. The children's failure to learn this task

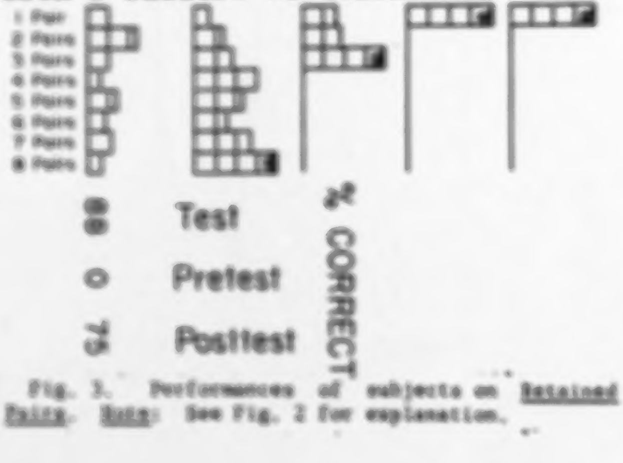


Fig. 3. Performances of subjects on *Retained Pairs*. *Note*: See Fig. 2 for explanation.

cannot be attributed to their inability to discriminate the graphemes or phonemes, since they passed the visual and auditory matching tests ( both of which closely approximated the training conditions). They failed to learn the correspondence.

Most of the children who were presented with a series of pairs failed on the first or second pair. Those who managed to get through two pairs tended to perform well on all subsequent pairs. However, three subjects reached criterion on more than two pairs before failing in a subsequent step, and two who completed eight successive pairs failed the test with the pairs scrambled -- even though they must have somehow detected the phoneme-grapheme relationship in order to discriminate between the members of each pair.

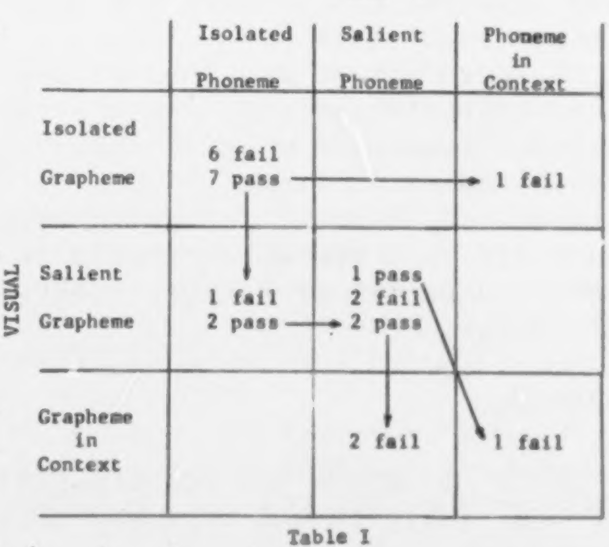
## EXPERIMENT II: MATCHING GRAPHEMES TO PHONEMES IN VARIOUS DEGREES OF ISOLATION

### Programs

We thought that a reasonable method of programing materials to teach this task would be to isolate the phonemes and graphemes, upon which the discrimination is based, from the remainder of the spoken and written words -- first teaching children to discriminate between the letters *s* and *l* in response to the phonemes /s/ and /l/ and then gradually introducing the other "irrelevant" auditory and visual elements. ( These elements are called "irrelevant" because the discrimination is not based upon them. This does not mean that they have no effect. Irrelevant phonemes affect the pronunciation of the relevant ones, and increasing the auditory and visual context requires the child to attend to the beginnings of the words. "Introducing irrelevant elements" is another way of talking about working up to a complex abstraction from simple instances.)

We decided to study three degrees of visual isolation: (a) the graphemes *s* and *l* alone, but placed in the position they would occupy if the remainder of the word were present; (b) words in which the "irrelevant" letters were partially erased, and (c) words as ordinarily written. Three comparable degrees of auditory isolation were: (a) the phonemes pronounced "alone" and accentuated ("sss" and "lll"); (b) words in which the phonemes were accentuated ("sssight", and "lllight"); and (c) words as ordinarily pronounced.

The possible combinations of these three degrees of auditory and visual isolation are shown in Table I. Upon entering the upper left hand cell of the Table (discriminating between the letters *s* and *l* in response to the phonemes /s/ and /l/) there are several alternative routes to the lower right hand cell (discriminating between written, "rhyming" words beginning with *s* and *l* in response to their spoken equivalents), and the routes may be traversed using a single pair of words or using several pairs of words at once. We composed materials to investigate several of the cells, using eight pairs of words in random order. Except for isolating the relevant phonic elements,



| | | AUDITORY | |
|---|---|---|---|
| | Isolated Phoneme | Salient Phoneme | Phoneme in Context |
| VISUAL Isolated Grapheme | 6 fail<br>7 pass → | → | → 1 fail |
| VISUAL Salient Grapheme | ↓<br>1 fail<br>2 pass → | 1 pass ↘<br>2 fail<br>→ 2 pass ↓ | |
| VISUAL Grapheme in Context | | ↓<br>2 fail | ↘ 1 fail |

Table I

the material was the same as that used in the Mixed Pairs program described earlier.

### Procedure

Two changes were made in the procedure at this time. (a) The equipment was modified so that the auditory stimuli were presented just before the visual stimuli appeared, rather than just afterward. Some children in the preceding experiments tended to respond to the written choices before the auditory sample was presented, and we thought that the new procedure might produce better stimulus control. (b) The loudspeaker was removed and the subjects were provided with earphones. This made it possible for one individual to run the experiment -- both giving instructions and pronouncing the sample words via a portable microphone.

Children (selected by means of the pre-test and the discrimination tests) were given the following instructions: "You are going to hear a sound and then you will see two different letters. Find the letter that goes with the sound." E pronounced a phoneme, pressed a hand switch, opening a shutter and exposing the choices, and waited for S to respond. E demonstrated if necessary, and when S had responded to four items E said, "Good, see if you can choose the right letter every time."

Children worked until they reached the criterion of six consecutive correct responses or until they had responded to 40 items. When criterion was reached the apparatus shut off and the experimenter inserted material of another "degree of isolation."

### Results

Table I shows the performances of 13 subjects who received this treatment. Nearly half of them failed to learn the isolated phoneme-grapheme correspondence. The arrows show what happened to the seven who succeeded in learning it: only one managed to reach the lower right hand cell and he failed to reach criterion there.

We were surprised to find the task of discriminating between isolated graphemes in response to isolated phonemes so difficult, and we began to doubt the adequacy of our experimental control. It was also clear from the results of this experiment that introducing the "irrelevant elements" would have to be done more delicately. The apparatus and procedure were therefore redesigned, and new programs were composed.

## EXPERIMENT III: PROGRAMING MATERIALS TO TEACH ISOLATED GRAPHEME-PHONEME MATCHING

### Apparatus and Procedure

A Tandberg Model 4F tape recorder was wired into the apparatus in order to present the auditory stimuli automatically. The operation of the new equipment was as follows: (a) When a subject was first seated *at the machine the screen was blank.* (b) On a signal from the experimenter the tape recorder played the first auditory sample. (c) The recorder stopped playing when it came to a pencil mark on the tape, and immediately the projector's shutter opened, displaying the choices. (d) A correct response caused the shutter to close, the projector changed slides, the tape recorder automatically played the next sample, and the shutter opened. This cycle was repeated until the end of the program.

The subject and the experimenter both listened to the tape recorder through earphones. The output from the upper track (4 ohms) was fed directly to two pairs of permoflex PDR-10 earphones equipped with sponge rubber pads. They adjusted the volume together at the start of the session.

The apparatus was also modified at this time to require a "ratio" of three responses to the correct window: A correct choice was recorded only when the proper window had been pressed three consecutive times. Incorrect responses reset the ratio and required the child to respond three times to the correct window again. The procedure had been found to improve performance in another matching task(Hively, 1964).

In the following experiments, subjects were not given visual and auditory discrimination tests. Those who failed to pass the pretest were introduced directly to the program.

## Pairing and Fading

A general method of programing discrimination materials is to make use of pre-established stimuli to elicit the required responses in the presence of the stimuli which are to come to control them. The old stimuli are then faded out or "vanished" in such a way as to transfer contol to the new stimuli. Skinner has made considerable use of the notion of "transfer of stimulus control" in his analysis of verbal behavior ( Skinner, 1957 ). He coined the term "vanishing" and described it as a programing technique (Skinner, 1958). Other writers (e.g., Glaser, 1962) have subsequently dealt with it. Fading and vanishing have only recently become the subject of experimentation (Terrace, 1961; Hively, 1962, 1964).

Our method was to first teach children to match the letters l and s to a visual sample (an identical letter). An auditory sample (the corresponding accentuated phoneme)was then paired with the visual sample, and the visual sample was faded out by covering it with Bourges Overlays (see Hively,1962, for a description of this techinque).

Program 1. Step 1 consisted of 10 visual matching items each of which was preceded by an auditory sample (each time the child made a correct choice the shutter closed, the slide changed, the auditory sample was presented,and the shutter opened displaying the choices and the visual sample). In Step 2 the visual sample faded out over the space of 8 items (two items in each "fading step"). Step 3 consisted of 10 items of the isolated grapheme-phoneme matching task.

No instructions were necessary for this program and those which follow. The children had been previously exposed to another visual matching task, and a short program (not shown in the figures) transformed the previous material into the letter matching task with which these programs began. The letter-matching section of the program constituted a kind of visual "pre-test." The auditory pre-test was not given.

Three of the six children who received Program 1 were out of control by Step 3. Since the same proportion failed the previous isolated grapheme - phoneme matching task (Table I),neither the changed apparatus and procedure nor the fading program appeared to have an appreciable effect upon the children's performances.

Program 2. In the preceding program, no contingency forced the children to "listen" to the auditory sample so long as the visual sample could be seen. In this program we tried to force auditory attention. Slides showing two pictures of familiar animals were seeded throughout the program. The auditory sample for these slides was a conventional imitation of one animal's characteristic sound: "moo", "cocka-doodle-doo", etc. In order to respond correctly to these "animal slides", the subject must attend to the auditory sample. We thought that seeding these items through the program might cause the children to listen more carefully. In any case, performance on these animal slides gives some indication of how carefully they were listening. Except for the addition of four animal frames in Step 2 and three in Step 3, this program was similar to the preceding one.

The performances of the six subjects who worked on this program were somewhat worse than those of the preceding group. Subject SGO made an error on an animal slide in Step 3, which suggests that he was out of control.

### Transforming

Instead of beginning with a visual matching task, pairing an auditory sample with the visual sample, and fading out the visual sample, one may begin with some pre-established auditory-visual correspondence and "transform" it into the grapheme-phoneme matching task. We tried teaching children to match pictures of animals to accentuated phonemes: a picture of a snake to the phoneme /s/ and a picture of a lizard to the phoneme /l/. Then the letters l and s were superimposed on the animal pictures and the pictures were faded out in three steps. Step 1 consisted of animal - phoneme matching (mixed with other animal slides). Steps 2 and 3 were fading steps. Step 4 was the grapheme-phoneme matching task.

In this program, too, performance deteriorated during the fading steps, and the final performances were not appreciably different from those which would have been expected from children who had received no program at all.

## PROGRAMING MATERIALS TO INTRODUCE IRRELEVANT ELEMENTS

For those children who did learn the isolated phoneme-grapheme correspondence, the problem was to introduce the remainder of the words. We investigated two programs: (a) introducing the irrelevant visual elements while holding the auditory elements constant, and (b) introducing the irrelevant auditory elements while holding the visual elements constant.

### Visual Program

Step 1 was the isolated grapheme-phoneme matching task. Step 2 consisted of three pairs of "erased" words presented first in "successive pairs" and then scrambled. Step 3 consisted of normally printed words, introduced by successive pairs and then all eight pairs were presented at random (the auditory sample was an isolated phoneme throughout).

This program proved to be effective. All four subjects correctly performed the task of discriminating pairs of "rhyming" written words in response to an accentuated initial consonant phoneme.

### Auditory Programs

Auditory Program 1. The same four subjects used in the preceding experiment were used on a program introducing irrelevant auditory elements. Step 1 consisted of the isolated grapheme-phoneme matching task again. Steps 2 and 3 represented two degrees of auditory isolation -- strongly accented initial consonants ("sssad", "lllad") and moderately accented ("ssad", "llad"). In Step 4 the words were pronounced normally. Eight pairs of words were presented in random order throughout, and the visual choices were always single letters.

Auditory Program 2. Step 1 was the isolated grapheme-phoneme task, Step 2 consisted of accentuated initial consonants followed by vowels alone: The final consonants of the words were not pronounced. Step 3 consisted of initial consonants and vowels, pronounced normally. Step 4 consisted of normally pronounced words. (Eight pairs of words were presented in random order throughout.) Eight animal slides (see Fading Program 2) were spaced through the program.

Auditory Program 3. In the above programs, the degree to which the phonemes were accentuated was difficult to control. A more adequate procedure was followed for program 3: words were pronounced with moderate emphasis on the initial consonants while moving the microphone varying distances away from the mouth. Step 1 was the isolated grapheme-phoneme task, Steps 2, 3 and 4 were different degrees of isolation, and Step 5 consisted of normally pronounced words (eight pairs in random sequence).

Of the 11 children who worked on the three auditory programs, only one learned the final task. It is surprising that the children found these programs so difficult.

## PROGRAMING MATERIALS TO TEACH GRAPHEME-PHONEME MATCHING IN SUCCESSIVE PAIRS OF WORDS

In the preceding programs we attempted, unsuccessfully, to teach the isolated phoneme-grapheme correspondence first and then to introduce the irrelevant visual and auditory elements. Another method is to apply the same techniques to individual pairs of words and to repeat the process with successive pairs. At first the auditory samples of a single rhyming pair (e.g., "llad" and "ssad") were paired with a letter matching task and the visual sample was faded out in a manner similar to that used in Fading Program 2. The child then was required to match letters to the words "sad" or "lad" spoken normally. Only then was the second rhyming pair introduced, first with the fading procedure then with the words spoken normally; and then a third pair was introduced and finally the three pairs were mixed.

Repeating the fading sequence with successive pairs of spoken words was no more effective as a method of teaching the auditory abstraction than the earlier programs were.

## SUMMARY AND DISCUSSION

Under the conditions of these experiments, matching the graphemes _s_ and _l_ to their corresponding phonemes when both appeared as initial consonants in pairs of "rhyming" words turned out to be a surprisingly difficult task. Several methods of programing materials to teach this behavior were tried and none proved to be effective.

In general there are three explanations for the inadequacy of a program of instruction:

a. The teaching situation may be so poorly arranged that the subject's behavior is never effectively shaped. Reinforcers may be inadequate, discriminative stimuli may be inadequately controlled, extraneous stimuli may not be sufficiently screened out, contingencies of reinforcement may not be precise enough to prevent unwanted behavior from being strengthened, and so on.

b. The program of instruction may be poorly constructed. The principles or techniques on which it is based, "pairing and fading" for example, may be ineffective under the given conditions, or specific parameters, such as the method and rate of fading, may have been badly chosen.

c. The criterion behavior as defined, e.g., matching graphemes to phonemes in the context of "rhyming" words, may simply be very difficult to teach.

The results of the present experiments will be briefly discussed in relation to each of these possibilities.

## Control of the Teaching Situation

The contingencies of reinforcement arranged in the present experiment were those which proved to be most effective in other studies ( Hively, 1964 ). Some notion of the degree of control to be expected may be obtained from those experiments. The auditory stimuli were inadequately controlled in the first part of the present experiments, but controlling them ( by means of the tape recorder ) had little effect.

Accurate performance was usually recovered when items from early steps in the programs were re-introduced during later steps. See, for example, the performance of subjects on animal slides seeded in later portions of the programs and on the repeated fading programs with individual pairs of words. This indicates that our teaching situation was reasonably well controlled. Nevertheless, the fact that we were able to teach the isolated grapheme-phoneme matching task to only half of the subjects, despite its apparent simplicity, makes it hard to be confident in the experimental control. The matter must be left to further experimentation.

## Principles Used in Constructing the Programs

*"Pairing and fading."* There is some reason to believe that fading out a visual sample in the presence of an auditory sample might not produce transfer of stimulus control. It is difficult to guarantee that the new stimuli are being observed in such a "cross-modal" situation.

*"Transforming."* It is less easy to see why transforming the visual stimuli from animal pictures to letters was not effective. In this case subjects are forced to observe both the "old" and the "new" stimuli in order to respond correctly during the fading program.

*Introducing irrelevant auditory and visual elements*. It is by no means clear why the visual program was effective and the auditory program was not. Both of these programs were "transforming" programs, in the sense that effective stimuli were gradually turned into ineffective stimuli without an obvious discontinuity in the process. Possibly the children had received more training in responding to individual letters in various contexts than in responding to individual phonemes, and they were therefore better prepared for the former program.

## Parameters

*Methods of fading*. The method of fading out visual stimuli by using *Bourges Overlays* may leave something to be desired. If one looks at the screen at the proper angle, the faded stimuli may be discriminated up to the last. This may account in part for the failure of the fading programs. A method which de-focussed or mutilated the stimuli ( see Israel, 1960 ) might have been more effective. However, it would be difficult to provide "smaller steps" in a program introducing irrelevant auditory elements than were arranged in Program 3. We would be hard pressed to explain its failure on those grounds.

*"Practice."* It is difficult to decide what number of "trials" or "items" to present at each step in a program. The optimum number may vary with the programing principle or technique being used, and with the size of the population of alternative instances. However, there is some evidence (Hively, 1964)that the point of diminishing returns is reached quite rapidly. We doubt that the effectiveness of the present programs would be greatly changed by increasing the num-

ber of items in each step.

Form of Criterion Behavior

The conclusion that a task is unteachable is, of course, a "null hypothesis." It cannot be proven, but it is tempting to accept it when many teaching methods are tried and nothing works. It might be argued that the present task is a peripheral one: one which children who have been trained to respond to the phonic correspondences in other situations are able to perform but which is difficult to teach directly. Other tasks, making use of the same words, which might lay claim to greater educational validity such as: writing in response to auditory stimuli, speaking words in response to printed stimuli, or even choosing among spoken alternatives in response to a printed sample.

Other programs of contrasts might be more effective in forcing discrimination of the phonic elements. Controlling the population of irrelevant elements appears to be an important consideration. Rather than dealing with a small set of relevant elements, such as *s* and *l*, and a large set of irrelevant elements such as is produced by all pairs of "rhyming" words, one might use a small set of words, such as *sad*, *sap*, *sid*, *sip*, *lad*, *lap*, *lid*, *lip*, and program materials to teach discrimination of the initial consonants, the vowels, and the final consonants as well -- all "irrelevant" elements in one set of discriminations being "relevant" in another.

## REFERENCES

Bloomfield, L. and Barnhart, C.L. *Let's read: a linguistic approach*. Detroit: Wayne University Press, 1961.

Daniels, J.C. and Diack, H. *Royal Road Readers*. Chatto and Windus, 1959.

Eaton, Winifred K. and James, Bertha F. *The Iroquois phonics series*. Syracuse, New York: Iroquois Publishing Co., 1956.

Glaser, R. Some research problems in automated instruction: instructional programming and subject-matter structure. In J. E. Coulson (Ed.), *Programmed learning and computer based instruction.* New York: John Wiley, 1962.

Hildreth, Gertrude. *Teaching reading*. New York: Henry Holt, 1958.

Hively, W. Programming stimuli in matching to sample. *J. exp. Anal. Behav.*, 1962, *5*, 279-298. [Included in this report.]

Hively, W. Parametric studies of a visual multiple-choice matching-to-sample program. [Included in this report.]

Israel, M. Variability blurred prompting in the analysis of paired associate learning. Unpublished doctoral dissertation, Harvard University, 1960.

Skinner, B. F. *Verbal Behavior*. New York: Appleton-Century-Crofts, 1957.

Skinner, B. F. Teaching Machines. *Science*, October 24, 1958, 128, 969-77.

Terrace, H. S. Discrimination learning with and without "errors". *J. exp. Anal. Behav.*, 1963, *6*, 1-27.

Thorndike, E. L. and Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College Columbia University, 1944.

PRECEDING PAGE BLANK-NOT FILMED

# VISUAL DISCRIMINATION OF ALPHABET LETTERS

## Helen M. Popp

THE PURPOSE of this study was to determine which pairs of alphabet letters are most often confused by prereading children and are, therefore, most likely to cause difficulty in initial reading instruction.* Although the confusions of the isolated letter-pairs may not necessarily be the same as those that would be encountered when letters are combined in a sequence, this study should help to indicate common specific confusions which effective instructional procedures might eliminate. It should also provide data for constructing a scale of items of relative difficulty to test the visual discrimination ability of youngsters beginning to learn to read.

Relevant studies include those by Hill (*9*), Davidson (*2*), and Smith (*13*). These are the only studies that were found which are concerned with the discrimination of single letters. Hill's test was concerned with specific categories of disorientation, i.e., "horizontal disorientations" (*b* and *d*), "vertical disorientations" (*p* and *b*), and "diagonal disorientations" (*p* and *d*). Therefore, the alternative choices in her matching task were limited to these categories. Davidson studied only the letters *d*, *e*, *n*, *q*, and *b* and concluded that reversals and inversions are the most frequent errors in letter discriminations. Her selection of choices was biased towards these types of confusions, however, so that her conclusions are not surprising. Both experimenters used a matching test with kindergarten children.

Smith tested first grade children on a letter matching task. Cards, each with a single letter printed on it, were presented one at a time in alphabetical order to an individual child, who was required to respond by finding an identical card from a scrambled array of the twenty-six letters. The study reports which single letters are most often responded to with an incorrect match, but no indication is given as to which pairs of letters are most frequently confused with one another.

Gibson *et al.* (*7*) studied the development of the ability to discriminate visually a set of letter-like forms, in children four through eight years of age, and related the types of errors made to certain critical features of letters. They hypothesized that certain *distinctive features* of letter patterns were responded to in the discrimination of letter-like forms. They suggested, for example, that "perspective transformations," "re-

---

*This report is based on research supported by a grant (719051) to the Harvard Department of Psychology from the United States Office of Education under Title VII. Subjects and space were made available through the cooperation of the School and University Program for Research and Development and the Angier Elementary School at Newton, Massachusetts. The author wishes to express gratitude to Dr. Wells Hively II for guidance and for making the research possible.

[Reprinted from THE READING TEACHER, January 1964.]

versals and rotations," and "changes of line to curve" in their letter-like forms were highly confusing to the younger (four-year-old) children because such transformations are not distinctive features for object identification in the children's past experience, and that the two latter types of errors drop markedly as the children learn to read, because the children develop a sensitivity to these same distinctive features. Errors on "perspective transformations" decreased only slightly, and the authors find this consistent with the fact that perspective changes are not critical for letter identification. Errors on "topological transformations" (a break and a close, as in *C* and *O*), however, were few, and the authors indicate that differences of this type are critical for object discrimination for the young child previous to his experiences in reading.

## Method

*Subjects.* Sixty-five children, age five years and one month to six years and one month, attending a public kindergarten, were divided into five groups of thirteen subjects each with an even distribution of chronological ages within groups. Testing was done in early February before the children had had formal training in letter matching or other prereading instruction.

*Apparatus.* The discrimination apparatus developed by, and more completely described by, Hively (*10*) was used. It consisted of a slide projector mounted behind a rear projection display panel that was sectioned into windows. A slide was projected showing a sample stimulus in the center section of the panel. The two alternative choice stimuli were displayed in the middle two of four windows just below the sample window. The child responded by pressing the window containing the stimulus he had chosen as matching the sample. A correct choice closed a switch and caused the projector to advance to the next slide. If the choice was incorrect, nothing happened and the subject would respond again. The subject sat at a low table facing the screen, his responses were recorded automatically, and the experimenter remained in an inconspicuous position in the experimental room.

*Materials.* Standard two-choice matching-to-sample items were constructed with each lower case letter of the alphabet (Century style print, projected ¼ inch in height) used as a correct choice in contrast to each of the other letters. There were, therefore, 25 items for each sample letter and a total of 650 items. A Gellerman Series (*6*) was used to determine the correct-choice position for each item constructed, and again to sequence the items, which were randomly selected from a pool of all the left-position-correct items and a pool of all the right-position-correct items. (A limit was imposed of no more than two successive items using the same sample stimulus.) Items were then divided into five sets of 130 each, and each group of 13 children was given a set of 130 items. The 650 letter items, therefore, were discriminated 13 times, and the 325 letter pairs were discriminated 26 times.

*Procedure*. All children had previously been exposed to a visual matching task using forms of different sizes, colors, and shapes. A short program consisting of fifteen slides was given in the discrimination apparatus, which transformed the colored forms into the letters *x* and *o*. These were decreased in size and made black in color so that the final discrimination was on items consisting of black letters *x* and *o*, projected to ⅛ inch in height. The effectiveness of this program was tested by requiring each subject to reach a criterion of seven successive correct responses in a matching task using *x* and *o* only. Two of the 65 children failed to reach this criterion within 20 items, and they were eliminated and replaced by children of the same age. Thirteen *x-o* items were seeded in with the 130 letter items presented to each child to check on experimental control and to give a "chance" distribution of errors. The letter discrimination test was given immediately following the pretraining in the same apparatus and without further instruction.

**Results**

The variance table for differences of error scores between the five groups is given below. The F-ratio was found to be 0.168, a value which is not significant; therefore, the differences between groups can be ignored.

| Source | Sum of Squares | df | Variance Estimate |
|---|---|---|---|
| Between | 47.80 | 4 | 11.94 |
| Within | 4253.00 | 60 | 70.88 |
| | 4300.80 | 64 | |

Table 1 indicates the number of errors made on each pair of letters.

TABLE 1
NUMBER OF ERRORS MADE ON EACH LETTER PAIR

[illegible]

Figure 1 is a frequency distribution of errors per pair. In order to estimate the probability that errors are due to discriminability and not to "chance," this distribution should be compared with a distribution of errors on pairs of letters which the subjects were known to be able to discriminate. Such a distribution was constructed from the error scores on the sixty-five *x-o* items. Each item with *x* as the sample was randomly paired with an item having *o* as the sample (one item was discarded), and error scores were tallied on these arbitrary "pairs." Figure 2 gives a frequency distribution of errors on these thirty-two pairs. The probability of a pair accumulating five or more errors, based on this distribution, is 1/32 or .03. All letter pairs with five or more errors are under-

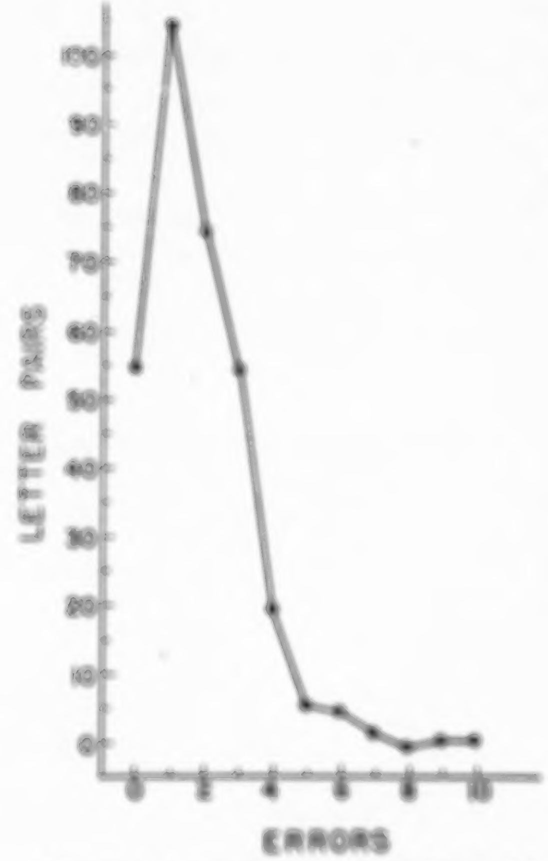


FIGURE 1. NUMBER OF LETTER PAIRS RECEIVING VARIOUS NUMBERS OF TOTAL ERRORS.

lined and listed separately in Table 1.

As expected, the pairs *b-d* and *p-q* are most confusing to these children, substantiating the results of Smith and Davidson. The subjects' error rate on the *x-o* slides was 6 per cent and only 7 per cent on the letter slides, indicating that the experimental control was good and also that the matching task was not difficult for these preschoolers. The range of errors per subject was from 0 to 34 on 143 items: only six subjects made over 20 errors.

A list indicating the relative difficulty of each letter by comparing percentages of error is found in Table 2. Smith's eleven most difficult (those with 10 per cent error or more) are also listed. Eight letters are common to both lists. It is interesting to speculate as to whether the differences in method, subjects, or possibly the differences in types of alphabet letters account for the discrepancies.

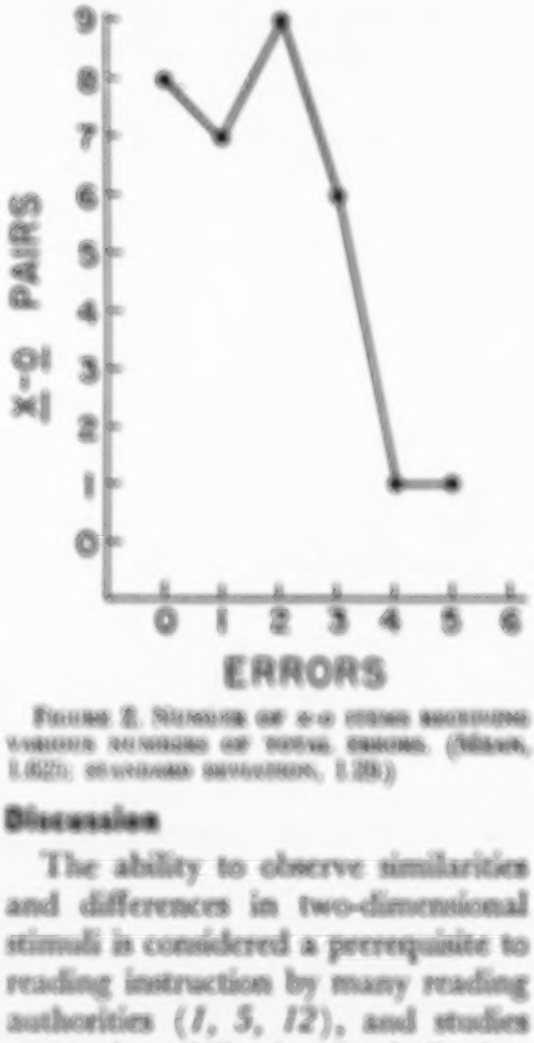


FIGURE 2. NUMBER OF X-O ITEMS RECEIVING VARIOUS NUMBERS OF TOTAL ERRORS. (MEAN, 1.62; STANDARD DEVIATION, 1.20)

## Discussion

The ability to observe similarities and differences in two-dimensional stimuli is considered a prerequisite to reading instruction by many reading authorities (*1, 5, 12*), and studies comparing ability in visual discrimination with later reading achievement most frequently report significant correlations (*3, 5, 8, 11*). At present, we can only speculate as to what effect specific training in improving competence in visual discrimination might have on later achievement in reading. There have

TABLE 2
GRAPHEMES WITH THE HIGHEST PERCENTAGE OF ERRORS

| Present Study | Per Cent, Errors | Smith (1928) Data | Per Cent, Errors |
|---|---|---|---|
| u | 10.00 | b | 34.1 |
| q | 09.69 | p | 28.1 |
| d | 08.92 | q | 22.5 |
| h | 08.77 | d | 21.6 |
| p | 08.77 | r | 14.1 |
| v | 08.00 | h | 13.3 |
| b | 07.85 | f | 13.3 |
| e | 07.69 | i | 12.5 |
| f | 07.54 | j | 12.5 |
| i | 07.54 | m | 12.5 |
| k | 07.54 | u | 10.8 |

been only a few studies in which any specific training has been attempted with children beginning to learn to read (the work of Goins in 1958, Muehl in 1960, 1961; and Staats, Staats and Schutz in 1962), and conclusions as to effectiveness are extremely tenuous. The various authors' conclusions are contradictory, and in many instances, the training procedures have been unsuccessful.

The present study was designed to measure the "confusability" of pairs of graphemes in a simple matching task for five-year-old children as a means of determining relative difficulty in discriminating such pairs.

The influence of similarity within the complete set of twenty-six letters can be assessed (rather than ranking individual letters as to "confusability" as Smith did). The results seem to substantiate the analysis as outlined thus far by Gibson (7) as confusions arise from reversals and rotations (*p-q*, *b-d*, *b-q*, *d-p*, *b-p*, *n-u*) and not from close and break transformations (*o-c*). Interpretation of the results as they relate to the features of letters which are critical for differentiation, and which may present difficulty, must wait for the forthcoming completed Gibson analysis.

Apparently some confusions in this study were also caused by a high degree of formal similarity (the proportion of similar or identical lines contained in both graphemes, *e.g.*, *i-l*, *k-y*, *h-n*).

It would seem that consideration of the distinctive features and of the formal similarity of letters should provide insight for training a specific skill in visual discrimination for letters. Remaining as a matter for further research is the problem of determining whether children trained to discriminate highly confusable letters on the basis of distinctive features and/or formal similarity will then be able to accurately discriminate all letters, and whether such training will have any effect on their later reading achievement.

(*Helen M. Popp is a doctoral candidate at the Harvard Graduate School of Education. She has been working for some time with Dr. Wells Hively on research concerned with basic programming questions, with one aspect that of the programming of phonic instruction for kindergarten children.*)

## References

1. Betts, E. A. *Foundations of Reading Instruction.* New York: American Book, 1954.
2. Davidson, Helen P. "A Study of the Confusing Letters *b*, *d*, *p*, and *q*," *Journal of Genetic Psychology*, 47 (1935), 458-468.
3. Fox, H. "The Relationship Between the Perception of Tachistoscopically Projected Images and Reading Readiness." Unpublished thesis, University of Indiana, 1953.

4. Gates, A. "A Further Evaluation of Reading Readiness Tests," *Elementary School Journal*, 40 (1940), 577-591.
5. Gates, A., and Boeker, Eloise. "A Study of Initial States in Reading by Preschool Children," *Teachers College Record*, 24 (1923), 469-488.
6. Gellerman, L. W. "Chance Orders of Alternating Stimuli in Visual Discrimination Experiments," *Journal of Genetic Psychology*, 42 (1933), 207-208.
7. Gibson, Eleanor, Gibson, J., Pick, Anne D., and Osser, H. "A Developmental Study of the Discrimination of Letter-Like Forms," *Journal of Comparative and Physiological Psychology*, 55 (1962), 897-906.
8. Henig, Max. "Predictive Value of Reading Readiness Tests and of Teachers' Forecasts," *Elementary School Journal* (1949), 41-46.
9. Hill, Mary, "A Study of the Process of Word Discrimination in Individuals Beginning to Read." Unpublished masters thesis, University of Iowa, 1935.
10. Hively, W. "Programming Stimuli in Matching to Sample," *Journal of the Experimental Analysis of Behavior* (1962), 279-298.
11. Kottmeyer, W. "Readiness for Reading," *Elementary English*, 24 (1947), 355-367.
12. McKee, P. *The Teaching of Reading in the Elementary School.* Boston, Houghton Mifflin, 1948.
13. Smith, Nila B. "Matching Ability as a Factor in First Grade Reading," *Journal of Educational Research*, 18 (1928), 560-571.
14. Smith, Nila B. "Reading Readiness" (a reprint from *Teacher's Guide for the First Year*). New York: Silver Burdett Co., 1936.

## COMMENT

Mrs. Popp's study provides some insight into the visual discrimination abilities of kindergarten children. The technique of investigating these abilities may seem unusual to persons in education, but the approach illustrates the procedures available from psychology which are applicable to educational problems. She carefully refrains from generalizing on the results of her study to the initial teaching of reading. She does, however, offer some interesting hypotheses for investigation.—T.C.

# SUMMARY

# REFLECTIONS ON A DECADE OF TEACHING MACHINES

### B. F. Skinner

To THE GENERAL PUBLIC, and to many educators as well, the nature and scope of teaching machines are by no means clear. There is an extraordinary need for more and better teaching, and any enterprise which may help to meet it will not be left to develop normally. The demand for information about teaching machines has been excessive. Articles and books have been published and lectures given; symposia have been arranged, and conferences and workshops have been held and courses taught. Those who have had anything useful to say have said it far too often, and those who have had nothing to say have been no more reticent.

Education is big business. Teaching machines were soon heralded as a growth industry, and fantastic predictions of the sales of programed texts were circulated. Devices have been sold as teaching machines which were not well built or designed with any understanding of their function or the practical exigencies of their use. No author was ever more warmly received by a publisher than the author of a programed text. Many programs, to be used either with machines or in textbook form, have been marketed without adequate evaluation.

## *Teachers and Devices*

The "mechanizing of education" has been taken literally in the sense of doing by machine what was formerly done by people. Some of the so-called computer-based teaching machines are designed simply to duplicate the behavior of teachers. To automate education with mechanical teachers is like automating banking with mechanical tellers and bookkeepers. What is needed in both cases is an analysis of the functions to be served, followed by the design of appropriate equipment. Nothing we now know about the learning process calls for very elaborate instrumentation.

Educational specialists have added to the confusion by trying to assimilate the principles upon which teaching machines are based to older theories of learning and teaching.

In the broadest sense, teaching machines are simply devices which make it possible to apply our technical knowledge of human behavior to the practical field of education *(10)*. Teaching is the expediting of learning. Students learn without teaching, but the teacher arranges conditions under which they learn more rapidly and effectively. In recent years, the experimental analysis of behavior has revealed many new facts about relevant conditions. The growing effectiveness of an experimental analysis is still not widely recognized, even within the behavioral sciences themselves, but the implications of some of its achievements

[Reprinted from TEACHERS COLLEGE RECORD, Vol. 65, No. 2, November 1963.]

for education can no longer be ignored.

An important condition is the relation between behavior and its consequences; learning occurs when behavior is "reinforced." The power of reinforcement is not easily appreciated by those who have not had firsthand experience in its use or have not at least seen some sort of experimental demonstration. Extensive changes in behavior can be brought about by arranging so-called contingencies of reinforcement. Various kinds of contingencies are concealed in the teacher's discussions with his students, in the books he gives them to read, in the charts and other materials he shows them, in the questions he asks them, and in the comments he makes on their answers. An experimental analysis clarifies these contingencies and suggests many improvements.

### *Shaping by Program*

An important contribution has been the so-called "programing" of knowledge and skills—the construction of carefully arranged sequences of contingencies leading to the terminal performances which are the object of education. The teacher begins with whatever behavior the student brings to the instructional situation; by selective reinforcement, he changes that behavior so that a given terminal performance is more and more closely approximated. Even with lower organisms, quite complex behaviors can be "shaped" in this way with surprising speed; the human organism is presumably far more sensitive. So important is the principle of programing that it is often regarded as the main contribution of the teaching-machine movement, but the experimental analysis of behavior has much more to contribute to a technology of education.

The direct contact which often exists between teacher and student favors the construction of programed sequences, and the teacher who understands the process can profit from the opportunity to improvise programs as he goes. Programs can be constructed in advance, however, which will successfully shape the behavior of most students without local modifications, and many of them can conveniently be mediated by mechanical devices. Laboratory studies have shown that contingencies emphasizing subtle properties of behavior can often be arranged *only* through instrumentation. There are potentially as many different kinds of teaching machines as there are kinds of contingencies of reinforcement.

Teaching machines which present material to the student and differentially reinforce his responses in well constructed programs differ in several ways from self-testing devices and self-scoring test forms, as well as from the training devices which have long been used by industry and the armed services. As Pressey pointed out many years ago *(8)*, a student will learn while taking a multiple-choice test if he is told immediately whether his answers are right or wrong. He learns not to give wrong answers again, and his right answers are strengthened. But testing has traditionally been distinguished from teaching for good reason. Before using a self-testing device, the student must already have studied the subject and, presumably, learned most of what he is to learn about it. Tests usually occupy only a small part of his time. Their main effect is motivational: A poor score induces him to study harder and possibly more effectively. Materials designed to be used in self-testing devices have recently been programed, but the contingencies which prevail during a test are not favorable to the shaping and maintaining of behavior.

Conventional training devices arrange conditions under which students learn, usually by simulating the conditions under which they eventually perform. Their original purpose was to prevent injury or waste during early stages of learning, but attention has recently been given to programing the actual behaviors they are designed to teach. To the extent that they expedite learning, they are teaching machines. Terminal performances have usually been selected for practical reasons, but a more promising possibility is the analysis and programing of basic motor and perceptual skills—a goal which should have an important place in any statement of educational policy.

In arranging contingencies of reinforcement, machines do many of the things teachers do; in that sense, they teach. The resulting instruction is not impersonal, however. A machine presents a program designed by someone who knew what was to be taught and could prepare an appropriate series of contingencies. It is most effective if used by a teacher who knows the student, has followed his progress, and can adapt available machines and materials to his needs. Instrumentation simply makes it possible for programer and teacher to provide conditions which maximally expedite learning. Instrumentation is thus secondary, but it is nevertheless inevitable if what is now known about behavior is to be used in an effective technology.

## *The New Pedagogy*

Any practical application of basic knowledge about teaching and learning is, of course, pedagogy. In the United States at least, the term is now discredited, but by emphasizing an analysis of learning processes, teaching machines and programed instruction have been responsible for some improvement in its status. The significance of the teaching machine movement can be indicated by noting the astonishing lack of interest which other proposals for the improvement of education show in the teaching process.

*Find better teachers.* In his *Talks to Teachers*, William James insisted that there was nothing wrong with the American school system which could not be corrected by "impregnating it with geniuses" *(7)*. It is an old formula: If you cannot solve a problem, find someone who can. If you do not know how to teach, find someone who knows or can find out for himself. But geniuses are in short supply, and good teachers do not come ready-made. Education would no doubt be improved if, as Conant *(3)* has repeatedly pointed out, good teachers who know and like the subjects they teach could be attracted and retained. But something more is needed. It is not true that "the two essentials of a good teacher are (a) enthusiasm and (b) thorough knowledge of and interest in his subject" *(5)*. A third essential is knowing how to teach.

*Emulate model schools.* Rickover's criticism of the present American school system is well-known *(9)*. His only important positive suggestion is to set up model schools, staffed by model teachers. The implication is that we already have, or at least can have for the asking, schools which need no improvement and whose methods can be widely copied. This is a dangerous assumption if it discourages further inquiry into instruction.

*Simplify what is to be learned.* Unsuccessful instruction is often blamed on refractory subject matters. Difficulties in teaching the verbal arts are often attributed to the inconsistencies and unnecessary complexities of a language. The pupil is taught manuscript handwriting because it more closely resembles printed forms. He is taught to spell only those words he is likely to use. Phonetic alphabets are devised to help him learn to read. It may be easier to teach such materials, but teaching itself is not thereby improved. Effective teaching would correct these pessimistic estimates of available instructional power.

*Reorganize what is to be learned.* The proper structuring of a subject matter is perhaps a part of pedagogy, but it can also serve as a mode of escape. Proposals for improving education by reorganizing what is to be learned usually contain an implicit assumption that students will automatically perceive and remember anything which has "good form"—a doctrine probably traceable to Gestalt psychology. Current revisions of high school curricula often seem to lean heavily on the belief that if what the student is to be taught has been "structured," he cannot help understanding and remembering it *(1)*. Other purposes of such revisions cannot be questioned: Materials should be up to date and well organized. But a high school presentation acceptable to a current physicist is no more easily taught or easily remembered than the out-of-date and erroneous material to be found in texts of a decade or more ago. Similarly, the accent of a native speaker encountered in a language laboratory is no more easily learned than a bad accent. No matter how well structured a subject matter may be, it must still be taught.

*Improve presentation.* Pedagogy can also be avoided if what is to be learned can be made memorable. Audio-visual devices are often recommended for this purpose. Many of their other purposes are easily defended. It is not always easy to bring the student into contact with the things he is to learn about. Words are easily imported into the classroom, and books, lectures, and discussions are therefore staples of education; but this is often an unfortunate bias. Audio-visual devices can enlarge the student's nonverbal experience. They can also serve to present material clearly and conveniently. Their use in attracting and holding the student's attention and in dramatizing a subject matter in such a way that it is almost automatically remembered must be questioned, however. It is especially tempting to turn to them for these purposes when the teacher does not use punitive methods to "make students study." But the result is not the same. When a student observes or attends to something in order to see it more clearly or remember it more effectively, his behavior must have been shaped and maintained by reinforcement. The temporal order was important. Certain reinforcing events must have occurred *after* the student looked at, read, and perhaps tested himself on the material. But when colored displays, attractive objects, filmed episodes, and other potentially reinforcing materials are used to attract attention, they must occur *before* the student engages in these activities. Nothing can reinforce a student for *paying* attention if it has already been used to *attract* his attention. Material which attracts attention fails to prepare the student to attend to material which is not interesting on its face, and material which is naturally memorable fails to prepare him to study and recall things which are not, in themselves, unforgettable. A well prepared instructional film may appear to be successful in arousing interest in a given subject, and parts of it may be remembered without effort, but it has not taught the student that a subject may *become* interesting when more closely examined or that intensive study of something which is likely to be overlooked may have reinforcing consequences.

*Multiply contacts between teacher and students.* Audio-visual devices, particularly when adapted to television, are also used to improve education by bringing one teacher into contact with an indefinitely large number of students. This can be done, of course, without analyzing how the teacher teaches, and it emphasizes a mode of communication which has two serious disadvantages: The teacher cannot see the effect he is having on his students, and large numbers of students must proceed at the same pace. Contributions to pedagogy may be made in designing programs for educational television, but the mere multiplication of contacts is not itself an improvement in teaching.

*Expand the educational system.* Inadequate education may be corrected by building more schools and recruiting more teachers so that the total quantity of education is increased, even though there is no change in efficiency.

*Raise standards.* Least effective in improving teaching are demands for higher standards. We may agree that students will be better educated when they learn more, but how are they to be induced to do so? Demands for higher standards usually come from critics who have least to offer in improving teaching itself.

The movement symbolized by the teaching machine differs from other proposals in two ways. It emphasizes the direct improvement of teaching on the principle that no enterprise can improve itself to the fullest extent without examining its basic processes. In the second place, it emphasizes the implementation of basic knowledge. If instructional practices violate many basic principles, it is only in part because these principles are not widely known. The teacher cannot put what he knows into practice in the classroom. Teaching machines and programed instruction constitute a direct attack on the problem of implementation. With appropriate administrative changes, they may bridge the gap between an effective pedagogical theory and actual practice.

### *Educational Goals*

An effective technology of teaching calls for a re-examination of educational objectives. What is the teacher's actual assignment? Educational policy is usually stated in traditional terms: The teacher is to "impart knowledge," "improve skills," "develop rational faculties," and so on. That education is best, says Dr. Hutchins *(6)*, which develops "intellectual power." The task of the teacher is to change certain inner processes or states. He is to improve the mind.

The role of the teacher in fostering mental prowess has a certain prestige. It has always been held superior to the role of the trainer of motor skills. And it has the great advantage of being almost invulnerable to criticism. In reply to the complaint that he has not produced observable results, the teacher of the mind can lay claim to invisible achievements. His students may not be able to read, but he has only been trying to make sure they wanted to learn. They may not be able to solve problems, but he has been teaching them simply to think creatively. They may be ignorant of specific facts, but he has been primarily concerned with their general interest in a field.

Traditional specifications of the goals of education have never told the teacher what to do upon a given occasion. No one knows how to alter a mental process or strengthen a mental power, and no one can be sure that he has done so when he has tried. There have been many good teachers who have supposed themselves to be working on the minds of their students, but their actual practices and the results of those practices can be analyzed in other ways. The well educated student is distinguished by certain characteristics. What are they, and how can they be produced? Perhaps we could answer by redefining traditional goals: Instead of imparting knowledge, we could undertake to bring about those changes in behavior which are said to be the conspicuous manifestations of knowledge, or we could set up the behavior which is the mark of a man possessing well developed rational power. But mentalistic formulations are warped by irrelevant historical accidents. The behavior of the educated student is much more effectively analyzed directly as such.

Contrary to frequent assertions, a behavioristic formulation of human behavior is not a crude positivism which rejects mental processes because they are not accessible to the scientific public *(12)*. It does not emphasize the rote learning of verbal responses. It does not neglect the complex systems of verbal behavior which are said to show that a student has had an idea, or developed a concept, or entertained a proposition. It does not ignore the behavior involved in the intellectual and ethical problem solving called "thinking." It does not overlook the value judgments said to be invoked when we decide to teach one thing rather than another or when we defend the time and effort given to education. It is merely an effective formulation of those activities of teacher and student which have always been the concern of educational specialists *(11)*.

Not all behavioristic theories of learning are relevant, however. A distinction is commonly drawn between learning and performance. Learning is said to be a change in some special part of the organism, possibly the nervous system, of which behavior is merely the external and often erratic sign. With modern techniques, however, behavior can be much more successfully studied and manipulated than any such inner system, even when inferences about the latter are drawn from the behavior with the help of sophisticated statistics. An analysis of learning which concentrates on the behavior applies most directly to a technology, for the task of the teacher is to bring about changes in the student's behavior. His methods are equally conspicuous: He makes changes in the environment. A teaching method is simply a way of arranging an environment which expedites learning.

### *Managing Contingencies*

Such a formulation is not easily assimilated to the traditional psychology of learning. The teacher may arrange contingencies of reinforcement to set up new *forms* of response, as in teaching handwriting and speech or nonverbal forms of behavior in the arts, crafts, and sports. He may arrange contingencies to bring responses under new kinds of *stimulus control*, as in teaching the student to read or draw from copy, or to behave effectively upon other kinds of occasions. Current instructional programs designed to fulfill such assignments are mainly verbal, but comparable contingencies generate nonverbal behavior, including perceptual and motor skills and various kinds of intellectual and ethical self-management.

A second kind of programing maintains the student's behavior in strength. The form of the response and the stimulus control may not change; the student is simply more likely to respond. Some relevant methods are traditionally discussed under the heading of motivation. For example, we can strengthen behavior by introducing new reinforcers or making old ones more effective, as in giving the student better reasons for getting an education. The experimental analysis of behavior suggests another important possibility: Schedule available reinforcers more effectively. Appropriate terminal schedules of reinforcement will maintain the student's interest, make him industrious and persevering, stimulate his curiosity, and so on; but less demanding schedules, carefully designed to maintain the behavior at every stage, must come first. The programing of schedules of reinforcement is a promising alternative to the aversive control which, in spite of repeated reforms, still prevails in educational practice.

In neglecting programing, teaching methods have merely followed the lead of the experimental psychology of learning, where the almost universal practice has been to submit an organism immediately to terminal contingencies of reinforcement (*13*). A maze or a discrimination problem, for example, is learned only if the subject acquires appropriate behavior before the behavior he brings to the experiment has extinguished. The intermediate contingencies are largely accidental. The differences in behavior and in rate of learning which appear under these conditions are often attributed to inherited differences in ability.

In maximizing the student's success, programed instruction differs from so-called trial-and-error learning where the student is said to learn from his mistakes. At best, he learns not to make mistakes again. A successful response may survive, but trial-and-error teaching makes little provision for actually strengthening it. The method seems inevitably committed to aversive control. For the same reason, programed instruction does not closely resemble teaching patterned on everyday communication. It is usually not enough simply to tell the student something or induce him to read a book; he must be told or must read and then be questioned. In this "tell-and-test" pattern, the test is not given to measure what he has learned, but to show him what he has not learned and thus induce him to listen and read more carefully in the future. A similar basically aversive pattern is widespread at the college level, where the instructor assigns material and then examines on it. The student may learn to read carefully, to make notes, to discover for himself how to study, and so on, because in doing so he avoids aversive consequences, but he has not necessarily been taught. Assigning-and-testing is not teaching. The aversive by-products, familiar to everyone in the field of education, can be avoided through the use of programed positive reinforcement.

Many facts and principles derived from the experimental analysis of behavior are relevant to the construction of effective programs leading to terminal contingencies. The facts and principles are often difficult, but they make up an indispensable armamentarium of the effective teacher and educational specialist. We have long since passed the point at which our basic knowledge of human behavior can be applied to education through the use of a few general principles.

## *Principle and Practice*

The difference between general principles and an effective technology can be seen in certain efforts to assimilate the principles of programed instruction to earlier theories. Programed instruction has, for example, been called "Socratic." It is true that Socrates proceeded by small steps and often led his students through an argument with a series of verbal prompts, but the example often cited to illustrate his method suggests that he was unaware of an important detail—namely, that prompts must eventually be "vanished" in order to put the student on his own. In the famous scene in the *Meno*, Socrates demonstrates his theory that learning is simply recollection by leading an uneducated slave boy through Pythagoras's Golden Theorem. The boy responds with the rather timid compliance to be expected under the circumstances and never without help. Although Socrates himself and some of those among his listeners who were already familiar with the theorem may have understood the proof better at the end of the scene, there is no evidence whatsoever that the boy understood it or could reconstruct it. In this example of Socratic instruction, at least, the student almost certainly learned nothing.[1]

A seventeenth-century anticipation of programed instruction has also been found in the work of Comenius, who advocated teaching in small steps, no step being too great for the student who was about to take it. Programing is sometimes described simply as breaking material into a large number of small pieces, arranged in a plausible genetic order. But size of step is not enough. Something must happen to help the student take each step, and something must happen as he takes it. An effective program is usually composed of small steps, but the whole story is not to be found in Comenius's philosophy of education.

Another venerable principle is that the student should not proceed until he has fully understood what he is to learn at a given stage. Several writers have quoted E. L. Thorndike (*15*) to this effect, who wrote in 1912,

> If, by a miracle of mechanical ingenuity, a book could be so arranged that only to him who had done what was directed on page one would page two become visible, and so on, much that now requires personal instruction could be managed by print.

In commenting on this passage, Finn and Perrin (*4*) have written, ". . . Here are the insights of a genius. History can very often teach us a lesson in humility—and it does here. The interesting question is: Why couldn't we see it then?" We might also ask, why couldn't Thorndike see it then? He remained active in education for at least 30 years, but he turned from this extraordinarily promising principle to another and—as it proved—less profitable approach to educational psychology.

It is always tempting to argue that earlier ideas would have been effective if people has only paid attention to them. But a good idea must be more than right. It must command attention; it must make its own way because of what it does. Education does not need principles which will improve education as soon as people observe them; it needs a technology so powerful that it cannot be ignored. No matter how insightful the anticipation of modern principles in earlier writers may seem to have been, something was lacking or education would be much further advanced. We are on the threshold of a technology which will be not only right but effective (*14*).

## *Criteria of Research*

A science of behavior makes its principal contribution to a technology of education through the analysis of useful contingencies of reinforcement. It also suggests a new kind of educational research. Thorndike never realized the potentialities of his early work on learning because he turned to the measurement of mental abilities and to matched-group comparisons of teaching practices. He pioneered in a kind of research which, with the encouragement offered by promising new statistical techniques, was to dominate educational psychology for decades. It led to a serious neglect of the process of instruction.

There are practical reasons why we want to know whether a given method or instruction is successful or whether it is more successful than another. We may want to know what changes it brings about in the student, possibly in addition to those it was designed to effect. The more reliable our answers to such questions, the better. But reliability is not enough. Correlations between test scores and significant differences between group means tell us less about the behavior of the student in the act of learning than results obtained when the investigator can manipulate variables and assess their effects in a manner characteristic of laboratory research. The practices evaluated in studies of groups of students have usually not been suggested by earlier research of a similar nature, but have been drawn from tradition, from the improvisations of skillful teachers, or from suggestions made by theorists working intuitively or with other kinds of facts. No matter how much they may have stimulated the insightful or inventive researcher, the evaluations have seldom led directly to the design of improved practices.

---

[1] The program of the *Meno* episode constructed by Cohen (2) is an improvement in that the student responds with less prompting.

The contrast between statistical evaluation and the experimental analysis of teaching has an illuminating parallel in the field of medicine. Various drugs, regimens, surgical procedures, and so on, must be examined with respect to a very practical question: Does the health of the patient improve? But "health" is only a general description of specific physiological processes, and "improvement" is, so to speak, merely a by-product of the changes in these processes induced by a given treatment. Medicine has reached the point where research on specific processes is a much more fertile source of new kinds of therapy than evaluations in terms of improvement in health. Similarly, in education, no matter how important improvement in the student's performance may be, it remains a by-product of specific changes in behavior resulting from the specific changes in the environment wrought by the teacher. Educational research patterned on an experimental analysis of behavior leads to a much better understanding of these basic processes. Research directed toward the behavior of the individual student has, of course, a long history, but it can still profit greatly from the support supplied by an experimental analysis of behavior.

This distinction explains why those concerned with experimental analyses of learning are not likely to take matched-group evaluations of teaching machines and programed instruction very seriously. It is not possible, of course, to evaluate either machines or programs *in general* because only specific instances can be tested, and available examples by no means represent all the possibilities; but even the evaluation of a given machine or program in the traditional manner may not give an accurate account of its effects. For example, those who are concerned with improvement are likely to test the student's capacity to give right answers. Being right has, of course, practical importance, but it is only one result of instruction. It is a doubtful measure of "knowledge" in any useful sense. We say that a student "knows the answer" if he can select it from an array of choices, but this does not mean that he could have given it without help. The right answer to one question does not imply right answers to all questions said to show the "possession of the same fact." Instructional programs are often criticized as repetitious or redundant when they are actually designed to put the student in possession of a number of different responses "expressing the same proposition." Whether such instruction is successful is not shown by any one right answer.

### *Correct or Educated?*

A preoccupation with correct answers has led to a common misunderstanding of programed materials. Since a sentence with a blank to be filled in by the student resembles a test item, it is often supposed that the response demanded by the blank is what is learned. In that case, a student could not be learning much because he may respond correctly in 19 out of 20 frames and must therefore already have known 95 per cent of the answers. The instruction which occurs as he completes an item comes from having responded to other parts of it. The extent of this instruction cannot be estimated from the fact that he is right 19 out of 20 times, either while pursuing a program *or on a subsequent test*. Nor will this statistic tell us whether other conditions are important. Is it most profitable for the student to execute the response by writing it out, by speaking it aloud, by speaking it silently, or by reading it in some other way? These procedures may or may not have different effects on a selected "right-answer" statistic, but no one statistic will cover all their effects.

Research in teaching must not, of course, lose sight of its main objective—to make education more effective. But improvement as such is a questionable dimension of the behavior of either teacher or student. Dimensions which are much more intimately related to the conditions the teacher arranges to expedite learning must be studied even though they do not contribute to improvement or contribute to it in a way which is not immediately obvious.

The changes in the behavior of the individual student brought about by manipulating the environment are usually immediate and specific. The results of statistical comparisons of group performances usually are not. From his study of the behavior of the individual student, the investigator gains a special kind of confidence. He usually knows what he has done to get one effect and what he must do to get another.

Confidence *in* education is another possible result of an effective technology of teaching. Competition between the various cultures of the world, warlike or friendly, is now an accepted fact, and the role played by education in strengthening and perpetuating a given way of life is clear. No field is in greater need of our most powerful intellectual resources. An effective educational technology based upon an experimental analysis will bring it support commensurate with its importance in the world today.

### REFERENCES

1. Bruner, J. S. *The process of education.* Cambridge: Harvard Univer. Press, 1960.
2. Cohen, I. S. Programed learning and the Socratic dialogue. *Amer. Psychologist*, 1962, *17*, 772-775.
3. Conant, J. B. *The education of American teachers.* New York: McGraw-Hill, 1963.
4. Finn, J. D., & Perrin, D. G. *Teaching machines and programmed learning: A survey of the industry, 1962.* Washington, D. C.: US Office of Education, 1962.
5. Helwig, J. Training of college teachers. *Sci.*, 1960, *132*, 845.
6. Hutchins, R. M. *On education.* Santa Barbara: Center for the Study of Democratic Institutions, 1963.
7. James, W. *Talks to teachers.* New York: Holt, 1899.
8. Pressey, S. J. A simple device for teaching, testing, and research in learning. *Sch. & Soc.*, 1926, *23*, 373-376.
9. Rickover, H. G. *Education and freedom.* New York: Dutton, 1959.
10. Skinner, B. F. The science of learning and the art of teaching. *Harvard educ. Rev.*, 1954, *24*, 86-97.
11. Skinner, B. F. Why we need teaching machines. *Harvard educ. Rev.*, 1961, *31*, 377-398.
12. Skinner, B. F. Behaviorism at fifty. *Sci.*, 1963, *140*, 951-958.
13. Skinner, B. F. Operant behavior. *Amer. Psychologist*, 1963, *18*, 503-515.
14. Skinner, B. F. *The technology of teaching* (in preparation).
15. Thorndike, E. L. *Education.* New York: Macmillan, 1912.

## Table 1

### Number and Monitoring Time in Minutes of Programs for Television Drama Content Analysis

| | | | % of total | |
|---|---|---|---|---|
| Program Type | n | time | n | time |
| Comedy | 56 | 1830 | 35.7 | 29.9 |
| Cartoon | 30 | 975 | 19.1 | 15.9 |
| Drama | 33 | 1645 | 21.0 | 26.8 |
| Western | 16 | 780 | 10.2 | 12.7 |
| Children's | 4 | 180 | 2.5 | 2.9 |
| Adventure | 11 | 390 | 7.0 | 6.4 |
| Police | 3 | 120 | 1.9 | 2.0 |
| Science Fiction | 2 | 90 | 1.2 | 1.5 |
| Mystery | 2 | 120 | 1.2 | 2.0 |
| Total | 157 | 6130 | | |

## Table 2

### Number of Programs Monitored and Monitoring Time Spent in Minutes by Analysts for Television Drama Content Analysis

| | Programs | | % | % |
|---|---|---|---|---|
| | n | Time | Programs Monitored | Time Monitored |
| Analyst$_1$ | 75 | 2710 | 47.8 | 44.2 |
| Analyst$_2$ | 30 | 1260 | 19.1 | 20.6 |
| Analyst$_3$ | 16 | 750 | 10.2 | 12.2 |
| Analyst$_4$ | 33 | 1260 | 21.0 | 20.6 |
| Analyst$_5$ | 3 | 150 | 2.0 | 2.4 |
| Total | 157 | 6130 | | |

## Table 3

### Television Programs Viewed: By Type, Time and Analyst

| Program | Type | Time | Analyst |
|---|---|---|---|
| Abbott & Costello | Comedy | 11:00 am Sat Oct 31 | 1 |
| Addams Family | Comedy | 7:30 pm Fri Oct 23 | 1 |
| | | Oct 30 | 2 |
| Alfred Hitchcock | Drama | 9:00 pm Mon Oct 19 | 3 |
| Alvin | Cartoons | 8:00 am Sat Nov 28 | 4 |
| Amos 'N' Andy | Comedy | 12:15 am Fri Oct 30/31 | 4 |
| Andy Griffith | Comedy | 10:00 am Thur Oct 15 | 1 |
| Andy's Gang | Children's | 8:00 am Sat Nov 21 | 4 |
| Annie Oakley | Comedy | 9:30 am Sat Oct 31 | 1 |
| Axel & Deputy Dawg | Cartoons | 4:30 pm Wed Oct 21 | 1 |
| Bachelor Father | Comedy | 3:00 pm Fri Oct 16 | 1 |
| Baileys of Balboa | Comedy | 8:30 pm Thur Nov 12 | 3 |
| Beany & Cecil | Comedy | 10:30 am Sat Oct 31 | 1 |
| Ben Casey | Drama | 9:00 pm Mon Oct 19 | 2 |
| Beverly Hillbillies | Comedy | 7:30 pm Wed Nov 18 | 2 |
| Bewitched | Comedy | 8:00 pm Thur Oct 15 | 2 |
| Bill Dana | Comedy | 7:30 pm Sun Oct 25 | 1 |
| | | Nov 8 | 4 |
| Bing Crosby | Comedy | 8:30 pm Mon Oct 19 | 2 |
| Bob Hope | Drama | 7:30 pm Fri Oct 30 | 1 |
| Bonanza | Western | 8:00 pm Sun Oct 18 | 2 |
| Broadside | Comedy | 7:30 pm Sun Nov 1 | 1 |
| Bullwinkle | Cartoons | 3:00 pm Sat Oct 17 | 1 |
| Buffalo Bill Jr. | Western | 8:30 am sat Nov 14 | 5 |
| Bugs Bunny | Cartoons | 11:00 am Sat Nov 7 | 1 |
| Burke's Law | Drama | 8:30 am Wed Oct 14 | 3 |
| Captain Kangeroo | Children's | 8:00 am Fri Nov 27 | 4 |
| Cara Williams | Comedy | 8:30 pm Wed Oct 14 | 2 |
| Car 54 | Comedy | 5:30 pm Sun Nov 8 | 4 |
| Casey & Roundhouse | Cartoons | 5:00 pm Tue Oct 20 | 1 |
| Casper | Cartoons | 10:00 am Sat Oct 17 | 1 |
| Circus Boy | Adventure | 12:00 pm Wed Oct 21 | 1 |
| Clancy & Axel | Cartoons | 7:00 am Fri Nov 27 | 4 |
| Clancy & Company | Cartoons | 5:00 pm Fri Oct 23 | 1 |
| Combat! | Drama | 6:30 pm Tue Oct 27 | 1 |
| Daniel Boone | Adventure | 6:30 pm Thur Dec 3 | 4 |
| Dave Lee & Pete | Cartoons | 3:30 pm Wed Oct 21 | 1 |
| Day In Court | Drama | 1:30 pm Mon Nov 2 | 1 |
| Death Valley Days | Drama | 9:30 pm Fri Oct 16 | 2 |
| Defenders | Drama | 9:00 pm Thur Oct 15 | 2 |

(Table cont. on next page)

| Program | Type | Time | Analyst |
|---|---|---|---|
| Dennis The Menace | Comedy | 10:00 am Sat Nov 7 | 1 |
| Desilu Playhouse | Drama | 11:00 am Fri Oct 16 | 1 |
| Dick Powell Theatre | Drama | 2:30 pm Sun Oct 18 | 2 |
| Dick Van Dyck | Comedy | 8:00 pm Wed Oct 14 | 2 |
| Divorce Court | Drama | 2:00 pm Sat Oct 31 | 1 |
| Dobie Gillis | Comedy | 6:00 pm Tue Oct 27 | 1 |
| Dr. Kildare | Drama | 7:30 pm Thur Nov 19 | 2 |
| Doctors/Nurses | Drama | 9:00 pm Tue Oct 20 | 2 |
| Donna Reed | Comedy | 7:00 pm Wed Nov 18 | 4 |
| 87th Precinct | Police | 3:30 pm Sun Oct 25 | 1 |
| | | Nov 8 | 4 |
| Ensign O'Toole | Comedy | 9:30 pm Wed Oct 14 | 2 |
| Farmer's Daughter | Comedy | 7:00 pm Fri Oct 23 | 1 |
| Father Knows Best | Comedy | 11:00 am Thur Oct 15 | 1 |
| Fireball XL-5 | Children's | 9:30 am Sat Oct 17 | 1 |
| Flintstones | Cartoons | 6:30 pm Thur Nov 26 | 4 |
| Flipper | Drama | 6:30 pm Sat Oct 31 | 1 |
| Forest Rangers | Adventure | 5:30 pm Sat Oct 31 | 1 |
| Fugitive | Drama | 9:00 pm Tue Oct 27 | 3 |
| Fury | Western | 10:30 am Sat Nov 7 | 1 |
| Gilligan's Island | Comedy | 7:30 pm Sat Oct 31 | 1 |
| Gomer Pyle-USMC | Comedy | 8:30 pm Fri Oct 16 | 3 |
| Gunsmoke | Western | 9:00 pm Sat Nov 7 | 2 |
| Hallmark Theatre | Comedy | 9:00 pm Sun Oct 18 | 2 |
| Harbor Command | Police | 12:30 pm Sun Nov 8 | 4 |
| Hazel | Comedy | 8:30 pm Thur Oct 29 | 2 |
| Hector Heathcote | Cartoons | 8:30 am Sat Nov 28 | 4 |
| Hoppity Hooper | Cartoons | 11:30 am Sat Nov 7 | 1 |
| Huckleberry Hound | Cartoons | 6:00 pm Wed Nov 11 | 1 |
| I Love Lucy | Comedy | 9:30 am Thur Oct 15 | 1 |
| Invisible Man | Drama | 4:30 pm Sun Nov 1 | 1 |
| Jack Benny | Comedy | 3:30 pm Fri Oct 16 | 1 |
| | | 7:30 pm Fri Nov 13 | 3 |
| Jackie Gleason | Comedy | 6:30 am Sat Nov 7 | 1 |
| Jetsons | Cartoons | 10:30 am Sat Oct 17 | 1 |
| Joey Bishop | Comedy | 8:30 pm Sun Oct 25 | 2 |
| Jonny Quest | Cartoons | 6:30 pm Fri Oct 23 | 1 |
| Kentucky Jones | Comedy | 7:30 pm Sat Oct 17 | 2 |
| King & Odie | Cartoons | 11:30 am Sat Oct 17 | 1 |
| Kraft Suspense Theatre | Drama | 9:00 pm Thur Oct 22 | 3 |
| Kuckla & Ollie | Cartoons | 10:40 am Fri Oct 16 | 1 |

(Table cont. on next page)

| **Program** | **Type** | **Time** | | | **Analyst** |
|---|---|---|---|---|---|
| Laramie | Western | 8:30 pm Tue | Oct | 20 | 3 |
| Lassie | Drama | 6:00 pm Sun | Oct | 18 | 1 |
| Leave It To Beaver | Comedy | 5:30 pm Tue | Nov | 3 | 1 |
| Linus The Lionhearted | Cartoons | 10:00 am Sat | Nov | 21 | 4 |
| Loretta Young | Drama | 1:00 pm Thur | Oct | 15 | 1 |
| Love that Bob! | Comedy | 12:00 pm Sat | Nov | 21 | 4 |
| Lucille Ball | Comedy | 8:00 pm Mon | Oct | 19 | 2 |
| Lunch With Casey | Cartoons | 12:45 pm Mon | Oct | 19 | 1 |
| Magilla Gorilla | Cartoons | 5:00 pm Fri | Nov | 27 | 4 |
| Make Room For Daddy | Comedy | 9:00 am Thur | Oct | 15 | 1 |
| Man From U.N.C.L.E. | Drama | 7:30 pm Tue | Nov | 17 | 2 |
| Many Happy Returns | Comedy | 8:30 pm Mon | Oct | 19 | 3 |
| Maverick | Western | 4:00 pm Fri | Oct | 16 | 1 |
| McCoys | Comedy | 10:30 am Thur | Oct | 15 | 1 |
| McHale's Navy | Comedy | 7:30 pm Tue | Oct | 27 | 1 |
| | | | Nov | 17 | 4 |
| Men Into Space | Adventure | 3:30 pm Sat | Nov | 7 | 1 |
| Men of Annapolis | Drama | 2:30 pm Sat | Nov | 1 | 1 |
| Mickey | Comedy | 8:00 pm Wed | Oct | 21 | 2 |
| Mighty Mouse | Cartoons | 9:30 am Sat | Oct | 31 | 1 |
| Mr. Broadway | Drama | 8:00 pm Sat | Oct | 17 | 3 |
| Mr. Ed | Comedy | 5:30 pm Sun | Oct | 18 | 1 |
| Mr. Lucky | Adventure | 2:30 pm Sat | Oct | 17 | 1 |
| Mr. Magoo | Cartoons | 7:00 pm Sat | Oct | 31 | 1 |
| Mr. Mayor | Children's | 7:00 am Sat | Nov | 21 | 4 |
| Mr. Novak | Drama | 6:30 pm Tue | Oct | 20 | 1 |
| Munsters | Comedy | 6:30 pm Thur | Nov | 19 | 4 |
| My Favorite Martian | Comedy | 6:30 pm Sun | Oct | 25 | 1 |
| | | | Nov | 8 | 4 |
| My Friend Flicka | Drama | 11:30 am Sat | Oct | 31 | 1 |
| My Little Margie | Comedy | 12:00 pm Sun | Nov | 15 | 4 |
| My Living Doll | Comedy | 8:30 pm Sun | Oct | 25 | 2 |
| My Three Sons | Comedy | 7:30 pm Thur | Nov | 19 | 2 |
| 90 Bristol Court | Comedy | 6:30 pm Mon | Nov | 9 | 4 |
| No Time For Sergeants | Comedy | 7:30 pm Mon | Oct | 26 | 4 |
| Official Detective | Police | 11:30 am Sun | Nov | 8 | 4 |
| One Step Beyond | Drama | 9:00 pm Sat | Oct | 17 | 3 |
| Outer Limits | Science Fiction | 6:30 pm Sat | Oct | 17 | 2 |
| Ozzie & Harriet | Comedy | 6:30 pm Wed | Nov | 18 | 4 |
| Patty Duke | Comedy | 7:00 pm Wed | Nov | 18 | 4 |
| People's Choice | Comedy | 12:30 pm Wed | Oct | 21 | 1 |
| Perry Mason | Mystery | 7:00 pm Thur | Nov | 26 | 4 |
| Peter Potamus | Cartoons | 5:00 pm Wed | Nov | 11 | 1 |
| Petticoat Junction | Comedy | 8:30 pm Tue | Oct | 27 | 3 |

(Table cont. on next page)

| Program | Type | Time | Analyst |
|---|---|---|---|
| Peyton Place | Drama | 8:30 pm Thur Oct 15 | 2 |
| Porky Pig | Cartoons | 9:30 am Sun Nov 8 | 4 |
| Profiles In Courage | Drama | 3:00 pm Sun Nov 8 | 1 |
| Quick Draw McGraw | Cartoons | 9:00 am Sat Nov 7 | 1 |
| Rawhide | Western | 6:30 pm Fri Oct 30 | 1 |
| Rebel | Western | 9:30 pm Sat Oct 17 | 3 |
| Red Skelton | Comedy | 7:30 pm Tue Nov 10 | 2 |
| Reporter | Drama | 9:00 pm Fri Oct 23 | 3 |
| Rifleman | Western | 6:00 pm Fri Oct 30 | 1 |
| Rogues | Drama | 9:00 pm Sun Nov 1 | 2 |
| Science Fiction Theatre | Science Fiction | 4:00 pm Sat Nov 7 | 1 |
| Sea Hunt | Adventure | 5:30 pm Sat Nov 14 | 4 |
| Sgt. Preston | Adventure | 10:00 am Sat Oct 31 | 1 |
| Sky King | Adventure | 11:00 am Sat Oct 17 | 1 |
| Slattery's People | Drama | 9:00 pm Mon Oct 26 | 3 |
| Soldiers of Fortune | Adventure | 11:30 am Sat Nov 21 | 4 |
| Studio 4 | Drama | 10:30 am Wed Nov 18 | 4 |
| Superman | Adventure | 4:30 pm Mon Oct 19 | 1 |
| Tennessee Tuxedo | Cartoons | 8:30 am Sat Nov 21 | 4 |
| This Is Alice | Comedy | 11:00 am Sun Nov 8 | 4 |
| Thriller | Mystery | 5:30 pm Sun Nov 29 | 5 |
| Top Cat | Cartoons | 12:00 pm Tue Oct 20 | 1 |
| Trailmaster | Western | 3:00 pm Mon Oct 19 | 1 |
| Twelve O'Clock High | Drama | 8:30 pm Fri Oct 16 | 2 |
| 26 Men | Western | 2:00 pm Sat Oct 17 | 1 |
| Tycoon | Comedy | 8:00 pm Tue Oct 20 | 2 |
| Underdog | Cartoons | 9:00 am Sat Oct 31 | 1 |
| Valentine's Day | Comedy | 8:00 pm Sat Oct 16 | 2 |
| Virginian | Western | 6:30 pm Wed Nov 11 | 1 |
| Voyage to the Bottom of the Sea | Adventure | 6:30 pm Mon Oct 26 | 4 |
| Wagon Train | Western | 6:30 pm Sun Nov 1 | 1 |
| Walt Disney's World | Western | 6:30 pm Sun Oct 18 | 1 |
| Wanted--Dead or Alive | Western | 10:00 pm Fri Oct 23 | 3 |
| Wendy & Me | Comedy | 8:00 pm Mon Oct 26 | 2 |
| West Point | Drama | 12:00 pm Sun Nov 8 | 4 |
| Woody Woodpecker | Cartoons | 6:00 pm Fri Nov 20 | 4 |
| Yancy Derringer | Western | 1:30 pm Sat Oct 17 | 1 |
| Yogi Bear | Cartoons | 6:00 pm Mon Nov 9 | 4 |

## Table 4

## Sample: Television Drama Content Analysis
## Code Sheet (first half page 1)

Coder's Name____________________ Time: Commenced__________

Concluded__________

Comic____________________ Total __________

Part 2

General Plot Outline:______________________________

______________________________

______________________________

______________________________

General Message:______________________________

______________________________

______________________________

______________________________

General Characters:______________________________

______________________________

______________________________

Type of Story: ____________________

Setting ____________________

Appeal: Sex ____________________

Age Group ____________________

Religion ____________________

Intellect ____________________

Personality ____________________

(Code sheet cont. on next page)

## Sample: Television Drama Content Analysis Code Sheet (second half page 1)

Revulsion: Sex ____________________

Age Group ____________________

Religion ____________________

Intellect ____________________

Personality ____________________

Additional Comments (if any): ____________________

(Code sheet cont. on next page)

## Sample: Television Drama Content Analysis
## (Code Sheet (first half page 2)

|  | Inferences |  |
|---|---|---|
| Description of Plot-Event | Comment on Character | Statement Re: Life-Reality |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Comments:

(Code sheet cont. on next page)

## Sample: Television Drama Content Analysis Code Sheet (second half page 2)

| Description of Plot-Event | Goals (ends) | Attitudes-Values |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

Comments: